
Biophysical and biochemical investigations of CARD-CARD interactions

Manuela Hess



The Francis Crick Institute
Mill Hill Laboratory
The Ridgway
London NW7 1AA, UK



Department of Structural and Molecular Biology
University College London

Thesis submitted to University College London for the degree of
Doctor of Philosophy

December 2015

Declaration of authenticity

The work presented in this thesis has been completed in the laboratory of Dr Katrin Rittinger at the Francis Crick Institute - Mill Hill Laboratory. I, Manuela Hess, confirm that this work is my own. Where information has been derived from other sources, this has been indicated in the thesis.

*[Diä Doktorarbeit isch mine Elterä und Grosselterä gwidmet, wo mich und mini Karriere
immer unterstützt händ.]*

Abstract

Innate and adaptive immunity are regulated by complex signalling pathways that require the assembly and activation of apoptotic and inflammatory signalling complexes, which ultimately induce gene expression and alter cellular behaviour. Key players in regulating the assembly of many of these signalling complexes are CARD-containing proteins that are members of the death domain superfamily. However, the molecular mechanisms that underline the function and regulation of CARDS remain unclear.

This project was aimed to investigate the biochemical and biophysical characteristics of CARD-CARD interactions to gain insights into the binding specificity and/or stoichiometry of CARD-mediated protein complexes and specifically, to elucidate if there is a common mode of interaction within this protein subfamily.

It was known that CARDS have a tendency to aggregate when purified individually, therefore all human CARDS were prepared for small scale expression and solubility screening with the aim to perform co-expression and pull-down studies. These screens identified the highly soluble, yet structurally and biophysically uncharacterised CARMA1-CARD. Consequently, one focus was set on the signalling complexes involving CARMA1 and CARD9, both of which bind BCL10 via CARD-CARD interactions.

Preliminary work had been performed in the group focusing on NOD2 and RIP2, therefore, these two proteins were taken as a second system for further investigations with the aim to describe the intramolecular interaction between the tandem CARDS of NOD2 and its intermolecular interaction with the CARD of RIP2 in molecular detail. Further insights into the binding surfaces involved in these interactions were gained using mutagenesis combined with biophysical assays.

To follow up on the newest development in the field concerning CARDS forming filamentous assemblies, this work tested both, RIP2- and BCL10-CARD for their ability to polymerise using fluorescence polymerisation assays. However, no polymerisation could be detected in this work, questioning the relevance of filament formation as a general behaviour of CARDS.

Table of contents

Declaration of authenticity	2
Abstract.....	4
Table of contents.....	5
List of figures and tables	10
Abbreviations	14
Acknowledgments	17
Chapter 1 - Introduction.....	18
1.1 The immune system - Innate and adaptive immunity	19
1.2 Transmembrane pattern recognition receptors	20
1.2.1 Toll-like receptors	20
1.2.1.1 TLR signalling pathways	22
1.2.2 C-type lectin receptors	24
1.2.2.1 Dectin-1 signalling pathway	25
1.3 Intracellular pattern recognition receptors	28
1.3.1 RIG-I like receptor signalling pathway	28
1.3.2 NOD-like receptor signalling pathways	31
1.3.2.1 The apoptosome	33
1.3.2.2 The NLRP3 inflammasome	34
1.3.2.3 The NOD2-RIP2 signalling pathway.....	35
1.4 Activation of transcription factors NF- κ B and IRF3/IRF7	39
1.4.1 Nuclear factor- <i>kappa</i> B (NF- κ B) - the canonical pathway	39
1.4.2 Interferon regulatory factors.....	40
1.5 Antigen receptors (AgR)	42
1.5.1 B-cell receptors (BCR)	42
1.5.2 T-cell receptors (TCR).....	43
1.5.3 Antigen receptor mediated activation and regulation of NF- κ B.....	43
1.5.3.1 CARMA1 activation	44
1.5.3.2 CBM-L complex formation and downstream signalling.....	46
1.6 The death domain superfamily	48
1.6.1 Complex formation within death domain subfamilies	50
1.6.2 The death domain (DD) subfamily	52
1.6.2.1 Intermolecular interaction	52
1.6.3 The death effector domain (DED) subfamily	54

1.6.3.1 MC159	55
1.6.3.2 Caspase-8	56
1.6.4 The pyrin domain (PYD) subfamily	58
1.6.5 The caspase recruitment domain (CARD) subfamily	58
1.6.5.1 Tandem CARDS and their intramolecular interactions	61
1.6.5.2 Intermolecular interactions	67
1.7 Project aims	71
Chapter 2 - Materials and methods	72
2.1 Materials.....	73
2.2 Molecular biology	73
2.2.1 Bioinformatics.....	73
2.2.2 Construction of expression plasmids.....	74
2.2.3 Polymerase chain reaction (PCR).....	75
2.2.4 Agarose gel electrophoresis.....	76
2.2.5 Determination of DNA concentration.....	76
2.2.6 Restriction enzyme digestion	77
2.2.7 Ligation independent cloning	77
2.2.8 Infusion cloning.....	79
2.2.9 Transformation of cloning products	79
2.2.10 Identification of positive plasmids and purification	80
2.2.11 Plasmid DNA purification, storage and sequencing	80
2.2.12 Site-directed mutagenesis	81
2.3 Protein expression and purification	82
2.3.1 Bacterial expression strains	82
2.3.2 Overexpression	83
2.3.2.1 Small-scale expression	83
2.3.2.2 Large-scale expression	83
2.3.3 Buffers and reducing agents.....	84
2.3.4 Small-scale solubility screen	84
2.3.5 Cell lysis and sample preparation	85
2.3.6 His-affinity chromatography.....	87
2.3.7 GST-affinity chromatography	87
2.3.7.1 Elution without tag removal	87
2.3.7.2 Elution including tag removal	88

2.3.7.3 GST-affinity purification of the modified tandem CARD construct of NOD2	88
2.3.8 Amylose-affinity chromatography	89
2.3.9 Ion exchange chromatography	89
2.3.10 Size exclusion chromatography	90
2.3.11 Protein refolding	90
2.3.11.1 Batch semi-refolding with 3 M urea	90
2.3.11.2 On-column refolding with 6 M guanidinium hydrochloride (GdmCl) ..	91
2.3.12 NMR sample preparation	91
2.3.13 SDS-PAGE electrophoresis	92
2.3.14 Determination of protein concentration	92
2.3.15 Concentration, verification and storage of protein samples	94
2.4 Biochemical and biophysical analysis	94
2.4.1 Co-expression and GST-pull downs	94
2.4.1.1 Small scale co-expression and purification screen of selected CARDS ...	95
2.4.1.2 NOD2-CARDab and RIP2-CARD	95
2.4.2 Quantitative analysis of protein-protein interactions	95
2.4.3 Nuclear magnetic resonance (NMR) spectroscopy	97
2.4.4 Isothermal titration calorimetry (ITC)	98
2.4.4.1 Determination of binding partners	99
2.4.4.2 Considerations and parameters in ITC	100
2.4.4.3 Materials and methods	102
2.4.5 Far-UV circular dichroism	103
2.4.5.1 General considerations	103
2.4.5.2 Materials and methods	105
2.4.6 Thermofluor	106
2.4.6.1 Materials and methods	106
2.4.7 SEC MALLS	106
2.4.7.1 Materials and methods	107
2.4.8 Fluorescence polarisation assays (FPA)	107
2.4.8.1 Instrumentation	108
2.4.8.1 Materials and methods	109
Chapter 3 - In search of soluble, interacting CARD pairs	111
3.1 Overview	112
3.2 Expression and solubility screens of all human CARDS	113

3.2.1 Construct design and cloning	113
3.2.2 Small scale expression and solubility screen	116
3.2.3 Co-expression and co-purification	125
3.2.4 Large scale test expression and purification of selected CARDS	126
3.3 Summary	130
Chapter 4 - Intra- and intermolecular CARD-CARD interactions of NOD2.....	131
4.1 Overview	132
4.2 The intramolecular interaction of the tandem CARDS of NOD2	133
4.2.1 NOD2 constructs	133
4.2.1.1 CARDa	133
4.2.1.2 CARDb	137
4.2.2 NMR titration experiments with the tryptophan CARDa mutants.....	139
4.2.3 Isothermal titration calorimetry with CARDa mutants	141
4.2.4 Circular dichroism to identify interacting CARDb residues	144
4.2.5 Isothermal titration calorimetry with CARDb mutants	148
4.3 Intermolecular CARD-CARD interactions of NOD2 and RIP2	150
4.3.1 Pull-down experiments with NOD2 tandem CARD mutants	151
4.3.1.1 Constructs used.....	151
4.3.1.2 Small scale co-expression and pull-down experiments	152
4.4 NOD2-CARDab and RIP2-CARD polymerisation study	155
4.4.1 Constructs, expression and purification	155
4.4.2 Fluorophore labelling.....	157
4.4.3 Biophysical analysis.....	158
4.4.3.1 UV spectrum.....	158
4.4.3.2 CD spectroscopy.....	158
4.4.3 Fluorescence polarisation assay with RIP2 and NOD2	159
4.5 Summary	163
Chapter 5 - Study of CARMA1, CARD9 and BCL10	164
5.1 Overview	165
5.2 CARMA1	166
5.2.1 CARD of CARMA1	166
5.2.1.1 Expression and purification.....	166
5.2.1.2 Crystallisation experiments.....	167
5.2.1.3 NMR spectroscopy	168

5.2.2 CARD-coiled-coil constructs of CARMA1	171
5.2.2.1 Construct design	171
5.2.2.2 SEC multi-angle laser light scattering.....	175
5.3 CARD9	177
5.3.1 CARD of CARD9	177
5.3.1.1 Constructs, expression and purification	177
5.3.1.2 CARD9-CARD aa 1-98: Thermofluor	179
5.3.1.3 Crystallisation experiments.....	182
5.3.2 CARD-coiled-coil constructs of CARD9	182
5.3.2.1 Constructs design.....	182
5.3.2.2 SEC multi-angle-laser-light scattering.....	185
5.4 BCL10	187
5.4.1 Constructs, expression and purification	187
5.4.2 Biophysical analysis of MBP-BCL10	189
5.5 Fluorescence polarisation assays with BCL10 and CARMA1.....	192
5.5.1 Constructs, expression and purification	193
5.5.2 Fluorophore labelling.....	195
5.5.3 Biophysical analysis.....	196
5.5.3.1 UV spectrum.....	196
5.5.3.2 CD spectroscopy.....	197
5.5.4 Fluorescence polarisation assay	197
5.6 Summary.....	202
Chapter 6 - Discussion and outlook	203
6.1 Overview.....	204
6.2 CARDS - expression and solubility screens	205
6.3 Intramolecular interaction between the tandem CARDS of NOD2.....	208
6.4 Intermolecular interaction of NOD2 and RIP2	212
6.5 CARMA1 and CARD9 - the CBM complex	215
6.6 Study of filamentous assemblies.....	217
6.7 Future directions	219
Bibliography	220
Appendix.....	229

List of figures and tables

Figure 1-1 Overview of innate immune signalling pathways.....	21
Figure 1-2 Model showing the TLR induced assembly of the MyDDosome.....	23
Figure 1-3 The functional domains of CARD9, BCL10 and MALT1.....	26
Figure 1-4 Structural model of the auto-inhibited and activated dsRNA-bound states..... of duck RIG-I.....	29
Figure 1-5 Examples of activating platforms.....	32
Figure 1-6 A structural insight into the apoptosome.....	33
Figure 1-7 The parallel downstream pathways of NOD2 signalling.....	38
Figure 1-8 Activation of NF- κ B and IRF transcription factors.....	41
Figure 1-9 Simplified model of early signalling events in AgR mediated NF- κ B..... activation.....	45
Figure 1-10 Cartoon of the domain structure of CARMA1, BCL10, and MALT1.....	46
Figure 1-11 Tertiary structures of the four death-fold subfamilies.....	50
Figure 1-12 Structure of Procaspase-9-CARD in complex with APAF1-CARD.....	51
Figure 1-13 Ternary structure of the MyDDosome.....	53
Figure 1-14 Structure of tandem death effector domains.....	57
Figure 1-15 Structure of RIG-I tandem CARDs.....	62
Figure 1-16 CD analysis of the NOD2 CARDa-CARDb complex.....	63
Figure 1-17 Solution structure of NOD2 CARDa.....	65
Figure 1-18 NMR saturation transfer difference spectra.....	66
Figure 1-19 Tryptophans of NOD2.....	66
Figure 1-20 EM model of a BCL10 filament.....	68
Figure 1-21 RIG-I-CARDab/MAVS-CARD complex.....	70
Figure 2-1 Ligation-independent cloning steps.....	78
Figure 2-2 UV absorption spectra of Trp, Tyr and Phe.....	93
Figure 2-3 Example of an ITC titration.....	101
Figure 2-4 Simulated binding isotherms at various c values.....	102
Figure 2-5 CD reference spectra.....	104
Figure 2.6 Schematic diagram of a general spectrofluorimeter.....	109
Figure 3-1 Multiple sequence alignment of all human CARDs.....	115

Figure 3-2 Small scale expression and purification.....	118
Figure 3-3 Small scale expression and solubility screen of His ₆ -tagged CARD.....	
constructs.....	119
Figure 3-4 Small scale expression and solubility screen of GST-tagged CARD constructs..	
using pET49b.....	121
Figure 3-5 Small scale expression and solubility screen of GST-tagged CARD constructs..	
using the new vector pGST.....	124
Figure 3-6 Small scale co-expression and co-purification.....	126
Figure 3-7 Purification of GST-CARMA1/-Iceberg.....	128
Figure 3-8 Purification of His ₆ -IPAF, -CARD6 and -ASC.....	129
Figure 4-1 Residues of MAVS-CARD essential for type 1, 2 and 3 interactions.....	135
Figure 4-2 Design of NOD2-CARDa mutants.....	136
Figure 4-3 MC159 tandem DED.....	139
Figure 4-4 Model of the tandem CARDS of NOD2.....	139
Figure 4-5 ITC analysis of NOD2-CARDa mutants.....	143
Figure 4-6 ITC analysis of additional NOD2-CARDa mutants.....	144
Figure 4-7 CD thermal unfolding of NOD2 tandem CARD mutants.....	146
Figure 4-8 Thermal unfolding of wild type constructs.....	147
Figure 4-9 CD thermal unfolding of NOD2 tandem CARD mutants after thrombin.....	
cleavage.....	149
Figure 4-10 ITC analysis of NOD2-CARDb mutants.....	151
Figure 4-11 Control gels of GST-pull-down experiments.....	154
Figure 4-12 GST-pulldowns: Effect of NOD2 CARDab mutants on RIP2 binding.....	155
Figure 4-13 GST-pulldowns with modified MBP-RIP2-CARD constructs.....	157
Figure 4-14 His ₆ -MBP-(3C)-RIP2-CARD C455A (aa 413-529) gel filtration purification.....	
after labelling.....	158
Figure 4-15 Far-UV CD spectra of labelled His ₆ -MBP-(3C)-RIP2 C455A (aa 413-529)...	160
Figure 4-16 Fluorescence spectra.....	161
Figure 4-17 Fluorescence polarization assay 1.....	162
Figure 4-18 Fluorescence polarization assay 2.....	162
Figure 4-19 SDS-PAGE analysis pre- and post-FPA.....	163
Figure 5-1 His ₆ -CARMA1-CARD purification.....	167

Figure 5-2 1D- ¹ H NMR spectrum of CARMA1-CARD.....	168
Figure 5-3 ¹ H- ¹⁵ N HSQC spectrum of CARMA1-CARD.....	69
Figure 5-4 ¹ H- ¹³ C- ¹⁵ N HNCO experiment of CARMA1-CARD.....	170
Figure 5-5 Prediction of coiled coil structures by MultiCoil.....	172
Figure 5-6 CARMA1-CARD-CC constructs.....	172
Figure 5-7 His ₆ -CARMA1 purification.....	174
Figure 5-8 SEC-MALLS analysis of CARMA1.....	176
Figure 5-9 His ₆ -CARD9-CARD purification.....	177
Figure 5-10 His ₆ -CARD9-CARD purification.....	178
Figure 5-11 Slice pH thermofluor assay with CARD9 aa 1-98.....	180
Figure 5-12 Solubility and stability thermofluor assay with CARD9 aa 1-98.....	181
Figure 5-13 Prediction of coiled coil structures by COILS and MultiCoil.....	183
Figure 5-14 CARD9-CARD-CC constructs.....	184
Figure 5-15 SEC-MALLS analysis of CARD9.....	186
Figure 5-16 His ₆ -BCL10-CARD purification.....	188
Figure 5-17 MBP-BCL10 aa 6-123 purification.....	189
Figure 5-18 Far-UV CD spectra of MBP and MBP-BCL10 aa 6-123.....	190
Figure 5-19 SEC-MALS analysis of MBP-BCL10 aa 6-123.....	191
Figure 5-20 His ₆ -MBP-(3C)-BCL10 C29/57A purification.....	194
Figure 5-21 His ₆ -MBP-(3C)-BCL10 C29/57A gel filtration purification.....	194
Figure 5-22 His ₆ -MBP-(3C)-BCL10 C29/57A gel filtration purification after labelling.....	195
Figure 5-23 UV spectra of the labelled His ₆ -MBP-(3C)-BCL10 C29/57A.....	196
Figure 5-24 Far-UV CD spectra of labelled His ₆ -MBP-(3C)-BCL10 C29/57A.....	197
Figure 5-25 Fluorescence spectra.....	198
Figure 5-26 Fluorescence polarization assay.....	199
Figure 5-27 Fluorescence polarization assay without glycerol.....	200
Figure 5-28 Fluorescence polarization assay with His ₆ -3C.....	201
Figure 5-29 SDS-PAGE analysis post FPA.....	201
Figure 6-1 Residues analysed via ITC mapped onto NOD2-CARDa.....	211
Figure 6-2 NOD-CARDa and RIP2 interaction.....	214

Table 1-1 Classification of CARD proteins.....	59
Table 1-2 CARD structures available as up to October 2015.....	60
Table 2-1 Typical PCR reaction mixture for cloning.....	75
Table 2-2 Typical PCR cycling protocol for cloning.....	76
Table 2-3 Typical digestion protocol.....	77
Table 2-4 Typical T4 treatment protocol.....	78
Table 2-5 Typical Infusion cloning reaction.....	79
Table 2-6 Transformation protocol using heat shock.....	80
Table 2-7 PCR reaction mixture for site-directed mutagenesis.....	81
Table 2-8 PCR cycling protocol for site-directed mutagenesis.....	81
Table 2-9 Lysis and purification buffers for small-scale expression and solubility.....	
screen.....	86
Table 2-10 Buffers for on-column refolding.....	91
Table 3-1 His ₆ -CARD constructs from the small scale expression and purification.....	
experiment.....	119
Table 3-2 GST-CARD constructs from the small scale expression and purification.....	
experiment.....	121
Table 3-3 Interaction partners of CARDS.....	124
Table 3-4: Small scale co-expression and purification.....	126
Table 4-1 Selected CARDa residues.....	134
Table 4-2 NOD2-CARDab mutants with mutations in CARDa.....	136
Table 4-3 NOD2-CARDab mutants with mutations in CARDb.....	139
Table 5-1 CARMA1-CARD-CC constructs.....	173
Table 5-2: Evaluation of 'Slice pH' thermofluor assays with CARD9 aa 1-98.....	180
Table 5-3: Evaluation of the 'Solubility and Stability' thermofluor assays with CARD9 aa	
1-98.....	181
Table 5-4 CARD9-CARD-CC constructs.....	185
Table 6-1 Soluble His ₆ -CARD constructs and their protein function.....	207

Abbreviations

Å	Angström
μL / μM	micro-litre / micro-molar
aa	Amino acid
ADP/ATP	Adenosine diphosphate/ adenosine triphosphate
AgR	Antigen receptor
APAF1	Apoptotic protease-activating factor 1
ASC	Apoptosis-associated speck-like protein containing a CARD
BCL10	B-cell lymphoma 10
BCR	B-cell receptor
BIR	Baculovirus inhibitor of apoptosis repeat
CARD	Caspase recruitment domain
CARMA1	CARD-containing MAGUK protein 1
CBM-L	CARMA1-BCL10-MALT1 complex in lymphoid cells
CBM-M	CARMA1-BCL10-MALT1 complex in myeloid cells
CC	Coiled-coil
CLR	C-type lectin receptors
CTD	C-terminal domain
Da	Dalton
DBD	DNA binding domain
DD	Death domain
DED	Death effector domain
DISC	Death-inducing Signalling Complex
DNA	Deoxyribonucleic acid
dsRNA	double-stranded ribonucleic acid
DTT	Dithiothreitol
FADD	FAS-associated protein with death domain
FPA	Fluorescence polarisation assay
FT	Flow through
GB1	Immunoglobulin binding domain of Streptococcal protein G
GdmCl	Guanidinium hydrochloride

GST	Glutathione S-transferase
ieDAP	γ -d-glutamyl-meso-diaminopimelic acid
IFN	Interferon
Ig	Immunoglobulin
I κ B	Inhibitor of kappa-B
IKK	I κ B kinase
IL-1 β /-18	Interleukin-1 β /-18
IPS-I	Interferon beta promoter stimulator protein 1
IPTG	Isopropyl β -D-1-thiogalactopyranoside
IRAK	IL-1 receptor-associated kinase
IRF	Interleukin regulatory factor
ITAM	Immunoreceptor tyrosine-based activation motifs
ITC	Isothermal titration calorimetry
JNK	C-jun N-terminal kinase
K _{A/D}	Association / dissociation constant
LB	Luria-Bertani broth
LGP2	Laboratory of genetics and physiology-2
LIC	Ligation independent cloning
LRR	Leucine rich repeat
MAGUK	Membrane-associated guanylate kinase
MALLS	Multi-angle laser-light scattering
MAPK	Mitogen-activated protein kinase
MAVS	Mitochondrial antiviral-signalling protein
MBP	Maltose binding protein
MDA5	Melanoma differentiation gene-5
MDP	N-Acetylmuramic acid L-Alanin-D-Isoglutamine
MHC	Major histocompatibility complex
M _w	Molecular weight
MyD88	Myeloid differentiation primary-response protein 88
NACHT	NAIP, CIITA, HET-E and TP1
NEMO	NF- κ B essential modulator
NF- κ B	Nuclear Factor-kappa- β

Ni-NTA	Nickel nitrilotriacetic acid
NLR	Nod-like receptor
NMR	Nuclear magnetic resonance
NOD	Nucleotide-binding oligomerisation domain
PAMP	Pathogen-associated molecular pattern
PCR	Polymerase chain reaction
PDB	Protein data bank
PDZ	Post synaptic density protein (PSD95)
pH	Power of hydrogen (-log10 of the hydrogen ion concentration)
pI	Isoelectric point
PKC β/θ	protein kinase C β/θ
PRR	Pattern recognition receptor
PYR	Pyrin domain
RIG-I	Retinoic acid inducible gene-I
RIP2	Receptor-interaction protein kinase
RLR	Retinoic-acid-inducible gene-I (RIG-I)-like receptors
RMSD	Root mean square deviation
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SH3	Sarcoma homology 3
SN	Supernatant
ssRNA	single-stranded ribonucleic acid
S/T-rich	Serine/Threonine rich
TAK1	Transforming-growth-factor- β -activated kinase 1
TBK1	TANK-binding kinase 1
TCEP	Tris(carboxyethyl)phosphine
TCR	T-cell receptor
TIR	Toll/IL-1R homology
TLR	Toll-like receptor
TNFR	Tumor necrosis factor receptor
TRAF6	TNF receptor associated factors 6
UV	Ultra-violet

Acknowledgments

It was hard... it was tough... but I survived four years working with members of the DEATH domain superfamily! So I got that goin' for me, which is nice.

Throughout the past few years I have asked myself the impossible question 'why?' many times. I may never have found an answer, but one thing was always crystal [ha...I wish] clear, that not in a single moment, did I ever regret the decision to take on this PhD position at the NIMR.

I first would like to thank my supervisor Dr Katrin Rittinger for giving me this incredible opportunity and for her guidance and support, even during the time when I lost the ability to walk. I am equally grateful to Dr Kovilen Sawmynaden and Dr Luigi Martino for their constant advice and patience. I would also like to express my thanks to current and past members of Molecular Structure and everyone who has made me laugh in the past four years; you have all made this place incredibly special. A huge thanks for Sarah¹, Sarah², Sarahⁿ⁺¹... Jan, Neil, David, Lizzi, Laura, Aylin, Laurence, Tom, Jonny, Marios and Liz for the never ending banter [waaayyy lads!] and giving me reasons to bake cakes.

Irene and Daniela, you two are my favourite geeks. Thanks for all the support from across the Channel.

The cottage people have also gained a special place in my heart; Clém, Martin, Alex, Jime, Charlie and Andy. There are no words which could summarise how much you all mean to me.

I would also like to thank my thesis committee members, Dr Paul Driscoll, Dr Luiz Pedro de Carvalho and Dr Steve Smerdon for their helpful discussions and putting up with my PowerPoint presentations, even if you had better things to do.

Chapter 1

Introduction

1.1 The immune system - Innate and adaptive immunity

Multicellular organisms are living in an environment surrounded by a high number of potentially harmful pathogens, such as bacteria, viruses, fungi and parasites. The first line of defence against invading microbial pathogens is the innate immune response; an evolutionary conserved mechanism shared by all multicellular organisms to rapidly initiate a host defence response within minutes. A secondary line of defence, a system developed just by vertebrates, is the adaptive immune response. The response is more delayed, but includes the benefit of an immunological memory to enable a faster secondary response in the future [1].

The innate immune response relies on a limited number of germline-encoded pattern-recognition receptors (PRRs) that belong to several distinct protein families. PRRs detect pathogen-associated molecular patterns (PAMPs). PAMPs are conserved products of microbial metabolism, which are essential for the survival of the microorganism, including viral nucleic acids, components of bacterial and fungal cell walls or flagellar proteins [2, 3].

Adaptive immune responses are carried out by two different classes of lymphocytes called B-cells and T-cells. Molecules capable of mounting an adaptive immune response are collectively called antibody generators, or simply antigens. The great diversity of receptor variants is generated by mechanisms involving rearrangement of receptor gene segments. Unlike the PRRs in the innate immune system, which detect molecular patterns, antigen-receptor binding involves recognition of specific features of biological molecules [1, 4].

Signal transduction plays a major role in both the innate and the adaptive immune system. The following sections will mainly focus on the innate immune system

with its PRR protein families followed by their downstream signalling pathways. An overview of selected pathways, which are relevant to this work can, be found in Figure 1-1. Signalling pathways relevant to acquired immunity will be briefly described and will focus mainly on the CARMA1-BCL10-MALT1 complex in lymphocytes.

1.2 Transmembrane pattern recognition receptors

Transmembrane PPRs enable signal transduction from the extracellular matrix to the cytoplasm for the activation of intracellular signalling pathways [5]. Transmembrane PPRs of innate immunity are foremost represented by the Toll-like receptors (TLRs) and the C-type-lectin receptors (CLRs). The following sections will concentrate on those signalling pathways that are important for the understanding of this thesis.

1.2.1 Toll-like receptors

Toll-like receptors are a family of evolutionarily conserved receptors that have a crucial role in early host defence against invading pathogens. The first TLR in human was reported in 1994 by Nomura *et al.* [6]. Since then, 10 TLRs have been identified which has driven forward the research of innate immunity. TLR members are type I transmembrane glycoproteins either expressed on the cell surface (TLR 1, 2, 4, 5, 6 and 10) or endosomal/lysosomal cell compartments (TLR 3, 7, 8 and 9) with molecular weights ranging between 90-115 kilodalton (kDa) [7].

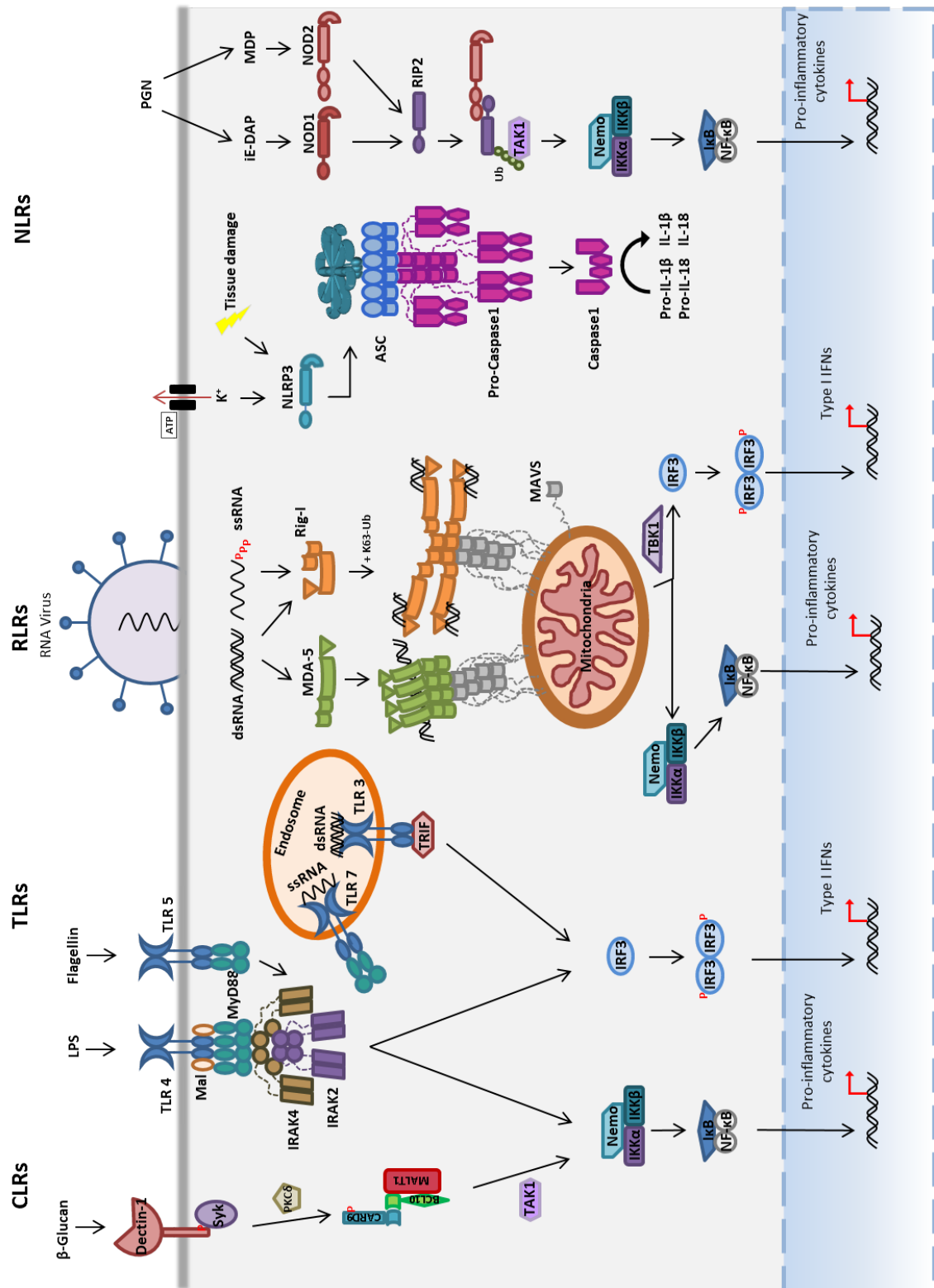


Figure 1-1 Overview of innate immune signalling pathways. Shown are the basic mechanisms after the activation of several signalling pathways of the innate immune system, divided into their pattern recognition receptor families. Details are provided in the text. Figure based on Yin *et al.* 2015, Kersse *et al.* 2011 and Ruland *et al.* 2008 [8-10].

TLRs are characterised by extracellular domains containing leucine-rich-repeat (LRR) motifs, thought to be directly involved in ligand binding, followed by a single, membrane-spanning α -helix and a cytoplasmic signalling domain called Toll/IL-1R homology (TIR) domain. The TIR domain is essential for the recruitment of downstream signalling molecules, ultimately leading to the induction of genes involved in the antimicrobial host defence response *i.e.* the production of inflammatory mediators [7].

1.2.1.1 TLR signalling pathways

Although LRRs are highly conserved, the human TLRs recognise a wide variety of PAMPs including viral nucleic acids and various types of bacterial and fungal ligands. Interestingly, individual TLRs interact with several structurally unrelated ligands such as lipopolysaccharides, RNA or proteins [11]. Upon ligand binding, TLRs are believed to dimerise and undergo conformational changes [12]. These conformational changes may affect the cytoplasmic TIR domain, required to recruit downstream signalling molecules with other TIR domains (primarily MyD88, MAL, TRIF and TRAM). The use of these adapter molecules varies between TLRs, which results in different signalling outcomes depending on the specific TLR activated. All TLRs except TLR3 recruit myeloid differentiation primary-response protein 88 (MyD88), which has an N-terminal death domain (DD), and a C-terminal TIR domain only separated by a short linker. MyD88 exists as homodimers in solution, where dimerisation is mediated by its death domain [13]. Ligand-induced dimerisation of TLRs leads to the recruitment of MyD88 to the receptors through homotypic TIR-TIR interactions. Although structures of both receptor and adaptor TIR domains are available (example protein data bank (PDB)

code: 2Z5V [14], 2J67 [15]), there is currently no structure of a TIR(receptor)-TIR(adaptor) complex, possibly due to difficulty in obtaining stable complexes in solution [16]. On the other hand, there is structural insight available into the recruitment of the downstream effector domain IL-1R-associated kinase (IRAK) by the dimeric MyD88 *via* DD-DD interactions [17] (Figure 1-2).

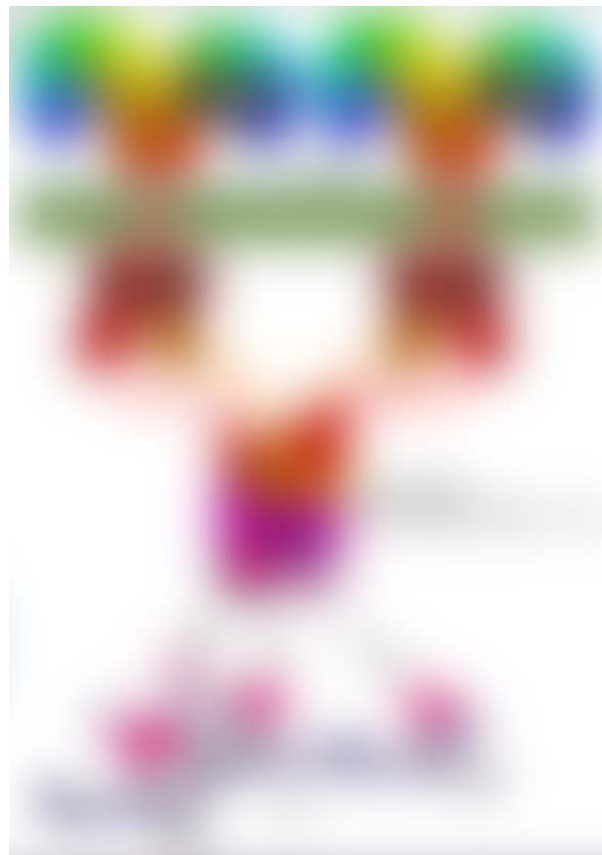


Figure 1-2 Model showing the TLR induced assembly of the MyDDosome. Ligand-induced dimerisation of TLRs leads to the recruitment of MyD88 (coloured as red or orange): Intracellular TLR-TIR domains (dark-red) form homotypic interactions with MyD88-TIRs. MyD88 linkers are shown as dashed lines. MyD88-DDs interacting with each other to recruit DDs of IRAK4 and IRAK1 or IRAK2 (pink and violet) for the assembly of a signalosome called MyDDosome. The MyDDosome interacts with TRAF6, an L-shaped E3-ligase, which can form a two dimensional lattice by forming dimers and trimers. TRAF6 enables the recruitment and activation of IKK and therefore the activation of NF- κ B. Figure reprinted from Berke *et al.* 2013 [18] with permission from John Wiley & Sons Inc.

IRAKs contain an N-terminal DD and a central serine/threonine-kinase domain. The crystal structure of a MyD88–IRAK4–IRAK2 death domain complex revealed an arrangement of six MyD88-DDs and four IRAK4-DD as well as four IRAK2-DDs into a left-handed helical tower. IRAKs phosphorylate and activate downstream substrates leading to the activation of NF- κ B. This MyD88–IRAK signalling platform was later named ‘the MyDDosome’. More structural details about the MyDDosome assembly will be provided in section 1.6.2.1.

DDs are members of the death domain superfamily, which includes death effector domains (DED), caspase recruitment domains (CARDs) and pyrin domains (PYDs). Effector domains in innate immune signalling are often members of this family. This protein superfamily will be explained in more details in section 1.6.

1.2.2 C-type lectin receptors

Other transmembrane PRRs are members of the C-type lectin receptors. The C-type lectins are the largest and most diverse of glycan binding receptors. Sugars are recognised by CLRs through their carbohydrate-recognition domain (CRD) in which ligand binding occurs in a Ca^{2+} -dependent manner [19]. Fungal PAMPs consist primarily of cell wall carbohydrate structures, therefore, CLRs are mostly known for their role in anti-fungal immunity [20]. For the purpose of this thesis, I will mainly focus on the signalling pathway of Dectin-1, a type II transmembrane enzyme that is expressed in myeloid cells including macrophages, dendritic cells and neutrophils [21].

1.2.2.1 Dectin-1 signalling pathway

The ectodomain of Dectin-1 recognises β 1-3 and β 1-6 linked β -glucans, which are categorised as fungal PAMPs [22]. Fungal recognition by Dectin-1 triggers phagocytosis and - more relevant here - it also induces signals mediated by CARD-containing proteins for the production of cytokines and chemokines as well as the activation of antigen presenting cells [10, 23]. Upon ligand binding, signalling is induced through tyrosine phosphorylation of the tyrosine-based activation motif (ITAM; YxxL) located within the cytoplasmic tail of these receptors. This is followed by the recruitment and activation of SYK kinase *via* its phosphotyrosine specific SH2 domain. This leads to the formation of a trimolecular complex composed of CARD9, BCL10, and MALT1, which will be referred from now on as the CBM-M complex, where 'M' stands for myeloid. The CBM-M complex ultimately leads to NF- κ B activation [10].

Role of CARD9 in the CBM-M signalling pathway

CARD9 is an adaptor protein containing an N-terminal CARD followed by a long coiled-coil domain, which functions as an oligomerisation domain. CARD9 plays a crucial role in the Dectin-1 signalling pathway to link signals from the ITAM receptor to the canonical NF- κ B pathway. CARD9 recruits the downstream adaptor protein B cell lymphoma 10 (BCL10) *via* a homotypic CARD-CARD interaction [24]. BCL10 is a 233 amino acid long protein containing an N-terminal CARD and a Ser/Thr-rich C-terminus of unknown function. The CARD9-BCL10 complex associates with the paracaspase MALT1 to complete the CBM signalosome [25]. An overview of the functional domains and known interactions is given in Figure 1-3. It has recently been shown that protein kinase C δ (PKC δ) is responsible for the phosphorylation of CARD9 at Thr231, a residue

located in the coiled-coil domain [26]. This phosphorylation was proven to be crucial to allow CARD9-BCL10 complex assembly. It is known, that the CMB-M is required for the activation of transforming-growth-factor- β -activated kinase 1 (TAK1) [26]. However, the molecular details of the signalosome activation remain unclear, as does the question of how exactly the CBM-M complex transduces signals to downstream effector molecules including TAK1, which finally leads to NF- κ B activation [10].

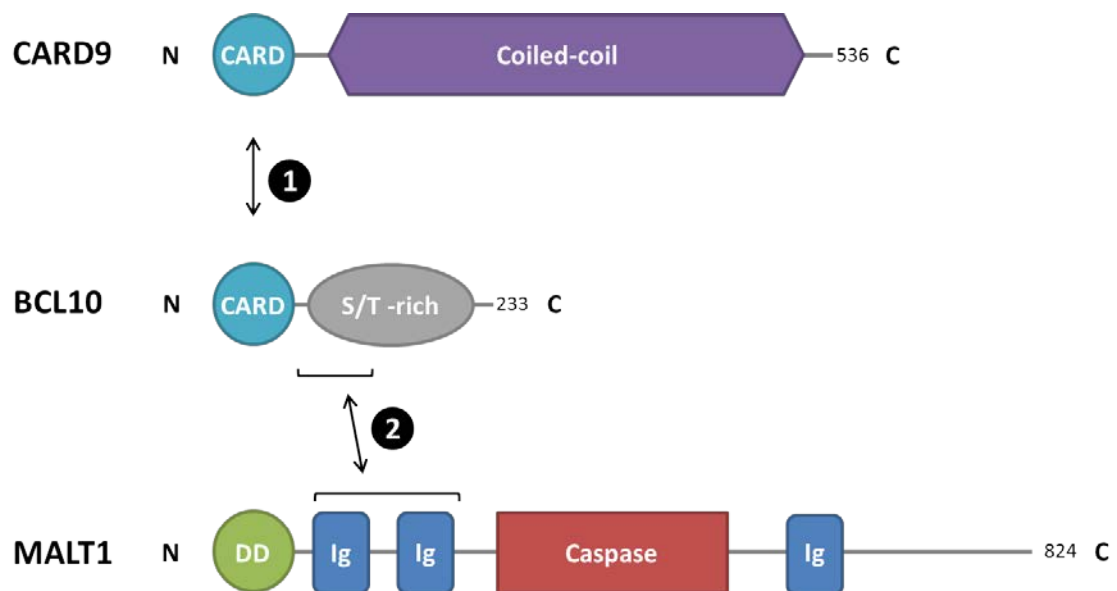


Figure 1-3 The functional domains of CARD9, BCL10 and MALT1. 1: In 2000, Bertin *et al.* showed *via* mammalian two-hybrid and co-immunoprecipitation assays that CARD9 associates with BCL10 through its CARD motif [24]. 2: In 2001, studies of Lucas *et al.* showed that MALT1 binds to BCL10 through its immunoglobulin domains [25]. CARD: caspase recruitment domain, DD: death domain, S/T -rich: Serine/Threonine rich region, Ig: immunoglobulin-like domain.

The role of CARD9 in this conserved CLR-triggered pathway has been shown to be essential in host protection. Patients who showed a point mutation from C to T in exon 6 at codon 295, resulting in a premature termination codon (Q295X) showed proneness to fungal infections as evidenced by chronic mucocutaneous candidiasis. *In vitro* reconstitution using CARD9^{-/-} cells suggested that the truncated CARD9 Q295X is a loss of function mutation as it failed to correct signalling defect observed in the CARD9 knock-out cells [27]. Further roles of CARD9 will be described in section 1.3.1 and 1.3.2.

1.3 Intracellular pattern recognition receptors

Once infectious agents have successfully invaded the cytosol, there are two families of cytoplasmic PRRs to detect pathogens and activate an innate immune response - the retinoic-acid-inducible gene-I (RIG-I)-like receptors (RLRs) and the NOD-like receptors (NLRs). RLRs recognise viral PAMPs such as dsRNA and NLRs can detect bacterial and endogenous ligands [28]. Recent structural insights into ligand detection and signal transduction revealed that intracellular PRRs assemble into multimeric protein complexes with their signalling adaptors upon ligand binding, as is the case for the transmembrane PRRs, which promotes recruitment of effector proteins and ultimately leads to inflammatory responses, principally through NF- κ B activation, type 1 interferons or Caspase-1 [8].

1.3.1 RIG-I like receptor signalling pathway

Members of the RLR family detect viral RNA present in the cytoplasm and initiate anti-viral immune responses [29]. The RLRs belong to the RNA helicase family and consist of only three members: retinoic acid inducible gene-I (RIG-I), melanoma differentiation gene-5 (MDA5) and laboratory of genetics and physiology-2 (LGP2) [30]. The proteins contain a central RNA helicase domain (DEXD/H-box helicase), which is flanked by a so-called C-terminal domain (CTD), while on their N-terminus, RIG-I and MDA5 contain tandem CARDs. It has been shown *in vitro* that RIG-I recognises short double stranded RNA (dsRNA) as well as single stranded RNA (ssRNA) with 5'-triphosphatate (5'-ppp) [31, 32]. Crystal structures revealed that RIG-I adopts an inactive state in absence of viral RNA, with its two CARDs interacting with the helicase domain [33]. RNA and ATP (adenosine triphosphate) binding induces a conformational

change to the helicase and CTD domain, which causes its two CARDs to disengage (Figure 1-4) [34, 35]. In the signalling pathway of RIG-I, both covalently attached and unanchored poly-ubiquitin chains were shown to play important roles: Back in 2007, Gack *et al.* [36] demonstrated *via in vivo* pull-down experiments and immunoblotting that the RIG-I tandem CARDs become ubiquitinated. They further identified the E3 ligase TRIM25 to interact with the N-terminal CARDs of RIG-I and catalyse their modification with Lys 63-linked ubiquitin chains, a process important for signalling activity. Subsequent studies indicated the importance of unanchored ubiquitin chains in RIG-I activation: The tandem CARDs of RIG-I assemble into a helical tetramer in complex with uncovalently bound Lys 63-linked di-ubiquitin molecules [37, 38].

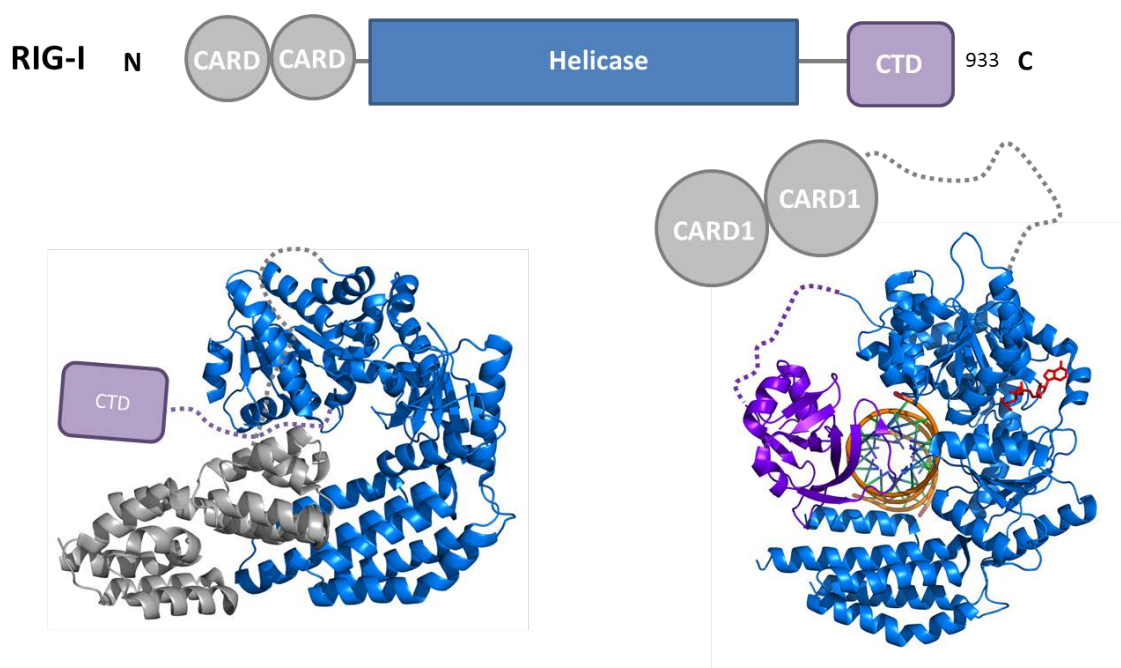


Figure 1-4 Structural model of the auto-inhibited and activated dsRNA-bound states of duck RIG-I (dRIG-I). Left: Model of the auto-inhibited dRIG-I with the tandem CARDs (grey), helicase domain (blue) and the flexible CTD (purple) (PDB: 4A2Q). Right: A model of the activated dRIG-I bound to dsRNA (orange) as well as the ATP transition state analogue ADP-AIF3 (red) (PDB: 4A36, 4A2X) with the tandem CARDs attached to a flexible linker but disengaged from the helicase domain. Figure based on Kowalinski *et al.* (2011) [33] and redrawn in PyMOL.

Following RIG-I activation, it is proposed that the RIG-I assembly nucleates the oligomerisation of the adaptor protein mitochondrial antiviral-signalling protein (MAVS, also called IPS-I) to promote formation of its active form. This nucleation leads to the polymerisation of MAVS-CARDs to form helical filaments similar to the death domains of MyD88 [39]. Interestingly, MAVS, which is located on the mitochondrial and peroxisomal membranes, was associated with self-propagating, prion-like properties to amplify signalling, drawing inactive monomers into functional aggregates [40]. In summary, the current model of RIG-I signalling as indicated in Figure 1-1 involves the oligomeric assembly of RIG-I upon RNA and ubiquitin binding / ubiquitination to initiate the oligomerisation and activation of MAVS. RLR signalling through MAVS oligomerisation ultimately leads to NF- κ B and IRF3 (Interferon regulatory factor 3) activation, inducing type-1 interferon production and other antiviral functions. More structural insights are given in section 1.6.5.2.

In contrast, comparison of the structures of MDA5 and RIG-I indicate significant differences in their conformations: MDA5, unlike RIG-I, doesn't seem to take on an auto-inhibited state in which the tandem CARDs are interacting with the helicase domain. Additionally, MDA5 has been shown to favour the binding of longer dsRNA (> 1 kbp) and to cooperatively assemble into ATP sensitive filaments on dsRNA [41, 42]. The MDA5-CTD was suggested to be required for cooperative filament formation rather than dsRNA binding [42]. This was further underlined with an EM study showing MDA5 forming a helical filament bound to dsRNA with the CTD forming contacts to adjacent MDA5 molecules [43]. An independent study showed similar results where the MDA5/dsRNA model was built using a 1:1 MDA5/dsRNA crystal structure and cross-linking experiments [44]. RIG-I has been also shown to assemble into helical

filaments and the current model suggests that RIG-I and MDA5 filament formation on dsRNA brings their CARDS into proximity to nucleate MAVS filament formation to initiate downstream signalling in an ubiquitin independent manner [42, 45].

1.3.2 NOD-like receptor signalling pathways

The NLR family contains 23 members in humans and is also known as NACHT-LRR, NOD-LRR or Caterpillar, all names which refer to the domains present in this family: Proteins of the NLR family are characterised by a tripartite domain structure. Similar to the TLRs, NLRs have a C-terminal receptor domain, which consists of a variable number of LRR motifs. Adjacent to this, NLRs have a nucleotide binding and ligand induced oligomerisation domain termed either NACHT (NAIP, CIITA, HET-E and TP1) or nucleotide-binding oligomerisation domain (NOD). Additionally, NLR family members have an N-terminal effector-binding domain, which is in most cases a member of the death domain superfamily that recruit and activate downstream effector molecules for signal transduction. NLRs are divided into the following subgroups according to their N-terminal domain: NLRA [A=acidic activation domain], NLRB; also called NAIP [B=baculovirus inhibitor of apoptosis repeat (BIR)], NLRC [C=Caspase recruiting domain] and NRLP [P=Pyrin domain] [46].

The ligand-induced oligomerisation of NLRs is followed by the recruitment of effector proteins to form large signalling assemblies that rely on homotypic interactions between members of the death domain family (Figure 1-5). Nowadays, NLRs are mostly known for their ability to initiate the formation of inflammasomes, supramolecular assemblies that mediate the activation of Caspase-1, which promotes maturation and secretion of the proinflammatory cytokines interleukin 1 β (IL-1 β) and

IL-18. The protein ‘apoptosis-associated speck-like protein containing a CARD’ (ASC) contains an N-terminal pyrin domain followed by a CARD and serves as the adaptor protein to bridge caspase recruitment to selected inflammasomes. Inflammasomes form in cells after sensing a diverse range of microbial, stress and damage signals. It is important that inflammasomes are tightly regulated as malfunctions can be linked to auto-inflammatory diseases (*e.g.* ulcerative colitis), Alzheimer’s disease and Crohn’s disease [47-49]. Therefore, regulation is complex and occurs at transcriptional and post-transcriptional levels.

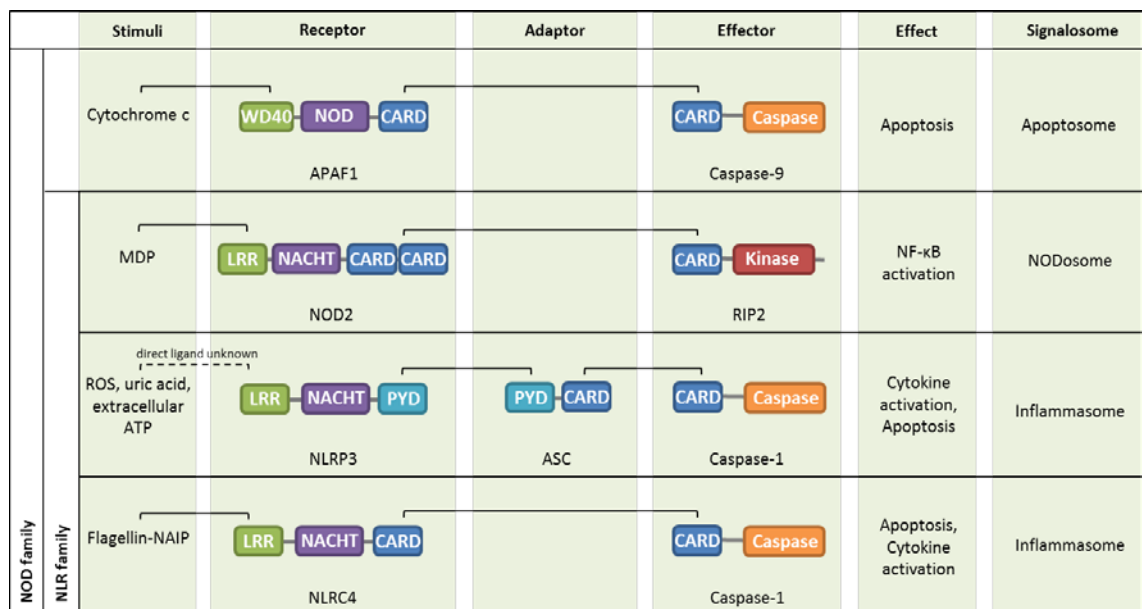


Figure 1-5 Examples of activating platforms. Shown are examples of signalling platforms with receptors of the NLR family and – for comparison – APAF1 from the wider nucleotide-oligomerisation domain (NOD) family. Proteins are divided into their role of sensors/receptors, adaptors or effectors. Figure based on Martinon *et al.* 2004 and Rathinam *et al.* 2012 [50, 51]. APAF1; Apoptotic protease-activating factor 1, NOD2; Nucleotide-binding oligomerization domain-containing protein 2, LRR; leucine-rich repeat, WD40; beta-transducin repeat, CARD; caspase-recruitment domain, PYD; pyrin domain, NACHT; NAIP/CIITA/HET-E/TP1 domain, ROS; reactive oxygen species, MDP; MurNAc-L-Ala-d-isoGln (peptidoglycan), Flagellin-NAIP; bacterial flagellin bound to baculoviral IAP repeat-containing protein 1 (NAIP).

1.3.2.1 The apoptosome

In the literature, NLRs are sometimes referred to as members of the wider NOD family. Another important member of the NOD family is the apoptotic protease activating factor 1 (APAF1). The main difference between NLRs and APAF1 is the type of ligand binding domain: whereas NLR proteins contain LRRs, APAF1 has WD-40 repeats (Figure 1-5). WD-40 is a structural motif of approximately 40 amino acids, often terminating in a tryptophan/aspartic acid (W-D) [52]. APAF1 exists in solution as an inactive monomer in which the WD-40 repeats are folded back onto the NOD thereby suppressing oligomerisation [53]. In presence of cytochrome c and ATP, APAF1 undergoes oligomerisation via its nucleotide binding domain. The EM structure of APAF1 bound to cytochrome c shows a heptameric disc, the so-called apoptosome (Figure 1-6) [54, 55].

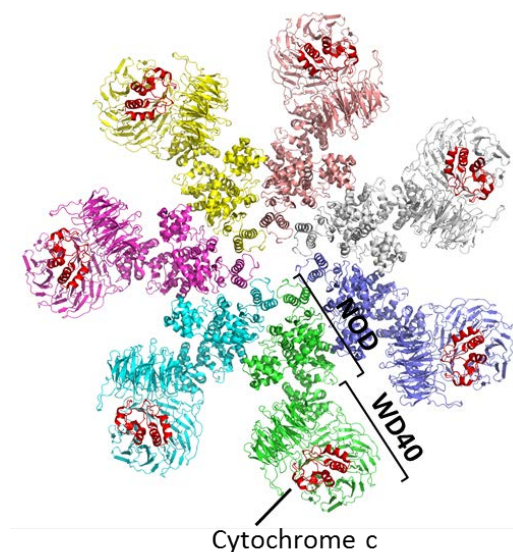


Figure 1-6 A structural insight into the apoptosome. The apoptosome is a multiprotein complex formed by APAF1 upon cytochrome c binding. In the shown cryo EM structure, the different APAF-1 molecules are highlighted in various colours and are lacking the CARD. WD-40 repeats of APAF1 bind to cytochrome c (red). In presence of ATP, the APAF-1-cytochrome-c complex forms a heptameric hub *via* its nucleotide oligomerisation domain (NOD). This Figure was prepared in PyMOL using PDB 3J3B.

The goal of APAF1 assembly is to form a platform that can recruit and activate the downstream effector Procaspase-9 to further activate caspases which ultimately mediate apoptosis. In the current model, APAF1 recruits Procaspase-9 via CARD-CARD interactions by assembling a ring-like CARD-platform. However, the precise mechanism is still unknown as the EM analysis only shows a low resolution disk sitting on the central heptameric APAF1 hub containing the APAF-1 and Procaspase-9 CARDS [54]. With no further structural insights, it is still under debate if activation of Procaspase-9 depends on the formation of a CARD-CARD disk.

1.3.2.2 The NLRP3 inflammasome

Out of the 23 human NLRs, six have been reported to assemble into inflammasomes: (NLRP1 [56], NLRP3 [57, 58], NLRP6 [48], NLRP7 [59], NLRP12 [60] and the NAIP/NLRC4 complex [61]). The NLRP3 inflammasome is the best-studied family member and for this reason, signalling *via* NLRP3 will be discussed in more detail (Figure 1-1). NLRP3 is activated by a wide range of stimuli from bacterial, viral, and fungal pathogens, as well as by endogenous danger signals indicating tissue damage (e.g. extracellular ATP, amyloid- β fibrils and uric acid crystals). However, with such a wide range of agents, it is believed that the interactions between the activators and the receptor are not direct [47]. Originally, the assembly model of NLRs was based on the heptameric hub formed by APAF1 (as seen in Figure 1-6), due to its similar domain organisation. Therefore, EM images of negatively stained specimens of NLRP1 revealing a low-resolution structure with a ring-like organisation were not surprising [56].

In addition to the initial receptor assembly of inflammasomes, in the last two years, more insight could be gained on ternary inflammasome complexes. In the NLRC3 pathway, it has been suggested, that there is a two-step nucleation-polymerisation mechanism: As shown in Figure 1-5, ASC is bridging NLRP3 and Caspase-1 by forming homotypic PYD-PYD with NLRP3 and CARD-CARD interactions with Caspase-1. Via EM and fluorescence polymerisation assays, it was shown that a sub-stoichiometric amount of NLRP3 nucleates ASC-PYD filament formation, which then further forms a platform for Caspase-1-CARD filaments. This leads to proximity driven dimerisation of Caspase-1, thereby inducing allosteric changes necessary for Caspase-1 auto-cleavage and activation. This allows Caspase-1 to process cytokine substrates pro-interleukin (IL)-1 β and pro-IL-18 [62]. Regulation of these cytokines is essential, whereas for example low concentration of IL-1 β is needed in local inflammation, higher concentrations can lead to fever and raised blood pressure and can in some cases be lethal [63].

1.3.2.3 The NOD2-RIP2 signalling pathway

NOD1 and NOD2

The first two NLRs identified were the nucleotide-binding and oligomerization-domain containing 1 and 2 (NOD1, NOD2) members of the NLRC subgroup [64]. Like the other NLRs described so far, NOD1 and NOD2 contain a tripartite domain structure with a C-terminal leucine-rich repeat (LRR) domain, a central NACHT domain and an N-terminal effector domain. Whereas this N-terminal effector region of NOD1 consists of one CARD motif, NOD2 contains two CARDS, referred to as tandem CARDS. Their LRR motif recognises fragments of bacterial peptidoglycan, more specifically, NOD1

recognises γ -d-glutamyl-meso-diaminopimelic acid (ieDAP) and NOD2 MurNAc L-Ala-d-isoGln (MDP) [65, 66]. Recent *in vitro* biochemical and biophysical studies have shown that the interaction with MDP happens in a direct fashion by immunoprecipitation and surface plasmon resonance assays [67, 68]. Similar to other NLRs, NOD1 and NOD2 undergo ligand-induced oligomerisation through their NACHT domains, which subsequently leads to the recruitment of receptor-interaction protein kinase 2 (RIP2) through CARD-CARD interactions [69].

RIP2

RIP2 is a 61 kDa serine/threonine kinase consisting of an N-terminal kinase domain and a C-terminal CARD (Figure 1-5). Serine/threonine kinases together with tyrosine kinases constitute one of the largest protein families in eukaryotes and phosphorylation is the most common post-translational modification in signal transduction [70]. Structural information on RIP2 is currently only available for the kinase domain in complex with an inhibitor (PDB: 4C8B) [71]. RIP2 mediates the activation of the NF- κ B [72, 73] and MAPK (mitogen-activated protein kinase) pathways [74] downstream of NOD1/NOD2.

Signalling pathway mediated by NOD2 and RIP2

There is currently no structural information available for activated forms of NOD2. Considering the similarity of the NOD2 domain structure with APAF1 and other NLRs, there is the possibility NOD2 may also form a disk-like hub as seen for the apoptosome. Another point of debate is whether both NOD2-CARDs are necessary for RIP2 recruitment. Studies available are contradictory and will be further discusses in

Chapter 6. Once the NOD2-RIP2 complex is formed, RIP2 becomes poly-ubiquitinated, currently assumed by the E3 ligase Pellino3 [75], to allow the recruitment of kinase TAK1 to the RIP2 kinase domain [76]. Signalling *via* NOD2 and RIP2 ultimately leads to NF- κ B activation to enhance transcription of inflammatory cytokines.

Regulation of NOD2 via CARD9 and association with Crohn's disease

It was discovered that invasion of epithelial cells by pathogenic bacteria responded not only with the activation of the transcription factor NF- κ B but also with the initiation of different MAPK pathways [77]. In 2007, Hsu *et al.* reported that this parallel pathway downstream of NOD2 is regulated by CARD9 (Figure 1-7) [78]. This is particularly interesting, because the genes for both of these proteins contain polymorphisms enhancing susceptibility to the inflammatory bowel condition Crohn's Disease in humans [79]. It was believed that CARD9 can form a tripartite complex with RIP2 and NOD2 *via* its CARD, which activates the JNK (C-jun N-terminal kinase) and p38 signalling pathways. These pathways upregulate expression of proinflammatory cytokines important for mounting appropriate immune responses. However, a recent biochemical and biophysical analysis showed that the interaction between NOD2 and CARD9 is mediated by two interaction sites in NOD2 outside the CARDs; one in the CARD-NACHT linker and one in the NACHT itself [80].

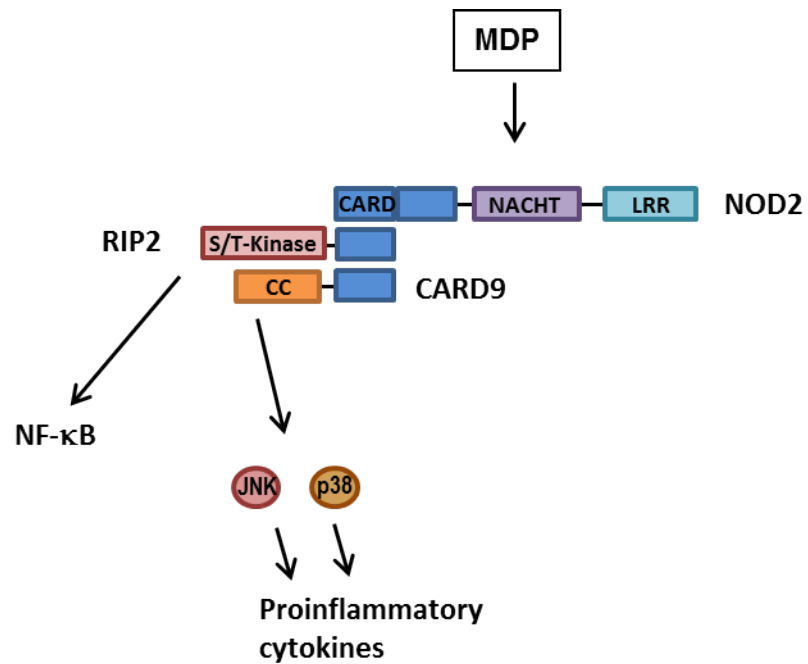


Figure 1-7 The parallel downstream pathways of NOD2 signalling. Whereas RIP2 is crucial for the activation of NF-κB signalling pathway, it is CARD9 which regulates the NOD2 induced activation of kinases p38 and JNK upon MDP recognition. LRR: Leucine-rich repeat, CC: coiled-coil, JNK: C-jun N-terminal kinase, MDP: MurNAc L-Ala-d-isoGln. Figure adapted from Roth and Ruland (2013) [81].

1.4 Activation of transcription factors NF- κ B and IRF3/IRF7

1.4.1 Nuclear factor-*kappa*B (NF- κ B) - the canonical pathway

NF- κ B is a transcription factor that forms either homo- or heterodimers. Subunits are members of the Rel-family, a family of structurally-related eukaryotic transcription factors. Mammalian NF- κ B includes five members; p50, p52, RelA, RelB, and c-Rel, which mediate activation of a large number of target genes by binding to an enhancer element called κ B. In unstimulated cells, NF- κ B dimers are retained in the cytoplasm by inhibitory proteins called inhibitors of κ B (I κ Bs) (Figure 1-8 A). Activation of NF- κ B occurs *via* two major pathways, the canonical and non-canonical pathway. The canonical pathway will be described here in more detail. As mentioned in previous paragraphs, transmembrane as well as intracellular PRR signalling pathways lead to the activation of the transcription factor NF- κ B. This happens usually through the recruitment of I κ B kinase (IKK) complex via adaptor proteins such as TAK1 kinases and/or TRAF (TNF receptor associated factors) ligases. The IKK consists of two kinases, IKK α and IKK β , and a regulatory subunit called IKK γ or NF- κ B essential modulator (NEMO). Upon recruitment, NEMO becomes ubiquitinated, which in turn activates the kinase activity of IKK α and IKK β . This allows the IKK α and IKK β subunits to phosphorylate I κ B α , which receives Lys48-linked ubiquitination by the E3 ubiquitin ligase TRCP, which means marking the subunit for proteasomal degradation. Subsequently, this unmasks the nuclear localisation sequence of NF- κ B and allows its translocation to the nucleus to initiate the expression of target genes [82-84].

1.4.2 Interferon regulatory factors

Interferon regulatory factors (IRFs) are, similar to NF- κ B, a family of transcription factors. In mammals, the IRFs have nine family members numbered as IRF1 to IRF9, which all of them contain a conserved N-terminal DNA binding domain (DBD) of about 120 amino acids. IRFs were originally thought to function mainly as crucial regulators of immune responses, however, recent studies suggest IRFs have additional critical functions such as in the regulation of oncogenesis or metabolism for example [85, 86]. In the regulation of innate immune responses, IRFs are acting, similar to NF- κ B, downstream of the PRRs and have the role of transmitting the signals of intracellular signalling cascades to chromatin to alter gene regulation.

Following detection of viral RNA via TLRs or RLRs, signalling cascades evoke in most cases type I IFN responses via activation of IRFs, particularly IRF3 and IRF7 (Figure 1-1). Activation of IRFs generally happens through phosphorylation by kinases such as TANK-binding kinase 1 (TBK1). TBK1 then phosphorylates IRF3 and IRF7, resulting in their homo- or hetero-dimerisation (Figure 1-8 B). This allows the translocation to the nucleus to induce expression of type I IFN genes [87, 88].

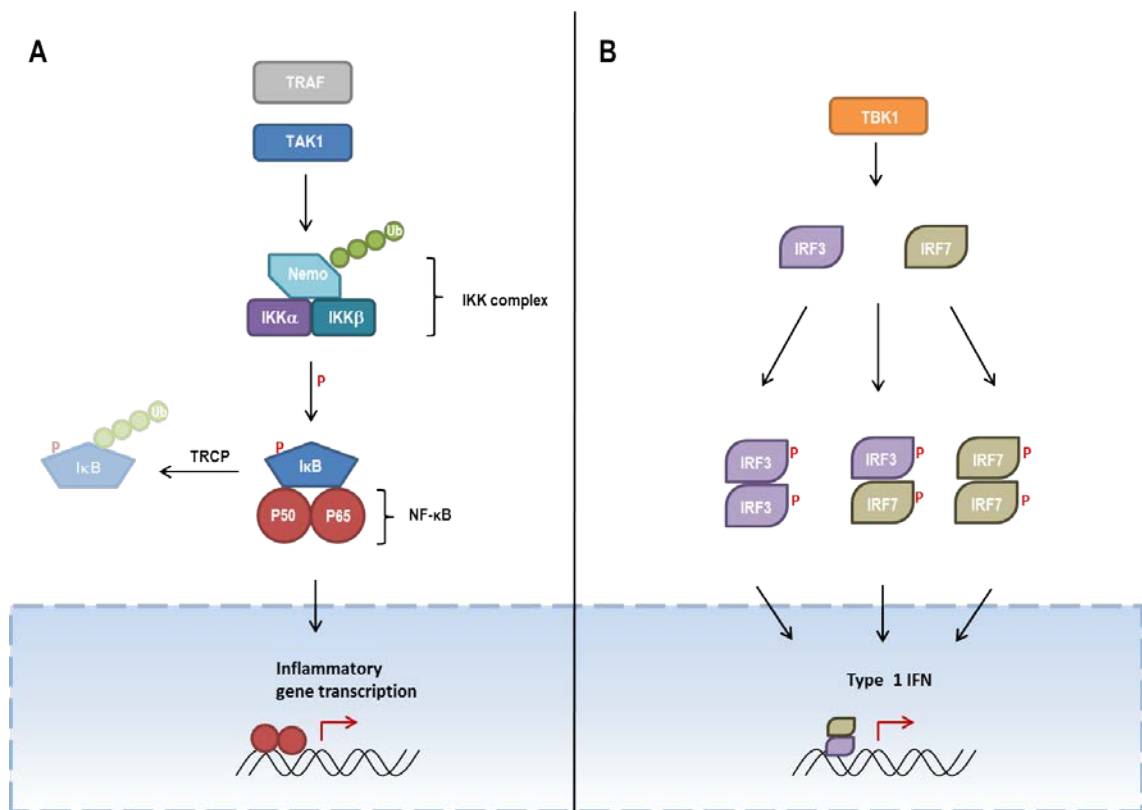


Figure 1-8 Activation of NF- κ B and IRF transcription factors. A: TAK1 kinases and/or TRAF ligases recruit the IKK complex, which is followed by the ubiquitination of NEMO. This activates the kinase activity of IKK α and IKK β , which phosphorylate I κ B α for proteasomal degradation. Subsequently, this allows NF- κ B to translocate into the cell nucleus to initiate the expression of target genes. B: Kinases such as TBK1 phosphorylate IRF3 and IRF7, resulting in their homo- or hetero-dimerisation. This allows the translocation to the nucleus to induce expression of type I IFN genes.

1.5 Antigen receptors (AgR)

The key players of the adaptive immune system are two different types of lymphocytes: B-cells and T-cells. For successful antigen detection and signal transduction to the cytoplasm, the cells require a number of antigen-specific cell surface receptors. The subunits for antigen binding and signal initiation differ between the two cell types [1].

1.5.1 B-cell receptors (BCR)

B-cells, originally formed in the bone marrow, are able to produce immunoglobulins, which can be either secreted into the blood stream or expressed as membrane-bound proteins functioning as cell surface antigen recognition units. As the membrane-bound immunoglobulin is lacking any intracellular signalling motive, it is linked to receptor-associated transmembrane accessory proteins: a heterodimer consisting of $Ig\alpha$ and $Ig\beta$ [1, 89]. $Ig\alpha$ and $Ig\beta$ have a cytoplasmic C-terminal region containing immune-receptor tyrosine-based activation motifs (ITAM), which are responsible for coupling signals from the extracellular matrix to the cell interior and activate downstream signalling pathways. ITAMs contain a conserved sequence of four amino acids in which a tyrosine is separated from a leucine or isoleucine by any two amino acids (YxxL/I). These sequence is generally repeated twice in the cytoplasmic domain of ITAM-containing proteins [90]. Activation of these intracellular signalling pathways alters gene expression, which ultimately leads to the neutralization and elimination of the antigen as well as B-cell differentiation to memory- or immunoglobulin secreting plasma cells.

1.5.2 T-cell receptors (TCR)

T-cells, developed in the thymus, contain antigen receptors, which are structurally similar to the BCRs. However, just a small number is able to recognise soluble antigens directly and therefore need antigen presenting cells, such as dendritic cells or macrophages. Antigen presenting cells load antigen peptides to major histocompatibility complex (MHC) molecules, which will then be transported to the cell surface and become accessible to TCRs [1, 91].

TCRs are composed of two disulfide-linked, immunoglobulin-like chains α and β ($\alpha\beta$ TCR) or γ and δ ($\gamma\delta$ TCR). Similar to BCRs, these chains lack a functional cytoplasmic tail for signal transduction and interact therefore with accessory molecules which contain ITAM motives; the CD3 chains (γ, δ, ϵ and ζ) [92, 93], which enable inducing downstream signalling cascades regulating the activity, proliferation and survival of the lymphocyte.

1.5.3 Antigen receptor mediated activation and regulation of NF- κ B

AgRs are key receptors to initiate signalling cascades, which regulate the proliferation and cell survival of lymphocytes. One way this is regulated, is through signalling cascades leading to NF- κ B activation. This signalling pathway is initiated by the activation of the SRC family of protein tyrosine kinases, which phosphorylate ITAMs within AgR signalling chains [94, 95]. Many factors downstream remain still unclear, but it has been shown that a trimolecular protein complex, involving the two CARD containing proteins CARMA1 and BCL10, is responsible for IKK complex activation. This complex, called CBM-L signalosome (CARMA1-BCL10-MALT1 in lymphoid cells), is

regulated by protein kinase C θ (PKC θ) in T-cells and PKC β in B-cells, respectively (Figure 1-9) [4, 96].

1.5.3.1 CARMA1 activation

CARD-containing MAGUK protein 1 (CARMA1), also referred as CARD11, is a 133 kDa scaffold protein and contains several domains, including CARD, coiled-coil, SH3, GUK, and PDZ domains (Figure 1-10) [97]. All of these domains apart from the PDZ domain have been shown to be required for the function of CARMA1 in mediating NF- κ B activation [98]. The structural arrangement of PDZ, SH3, and GUK domains is called the MAGUK (Membrane-associated guanylate kinase) domain. MAGUK family proteins bind directly to the cytoplasmic termini of transmembrane proteins and to other signal transduction proteins [99]. Additionally, in both, T and B cells, clustering of CARMA1 through its SH3-GUK domain is essential for NF- κ B activation [100]. CARMA1 appears to be regulated by a range of phosphorylation events; It contains phosphorylation sites within the linker region between the coiled coil and the PDZ domains, which are targets of PKC θ and PKC β , suggesting CARMA1 to adopt an inactive conformation that is stabilized by an intramolecular interaction between the CARD and the linker region [101, 102]. Hence, it is believed that these phosphorylation events trigger a conformational change, which enables BCL10-MALT1 recruitment, as the CARD domain and the coiled-coil of CARMA1 become accessible [102].

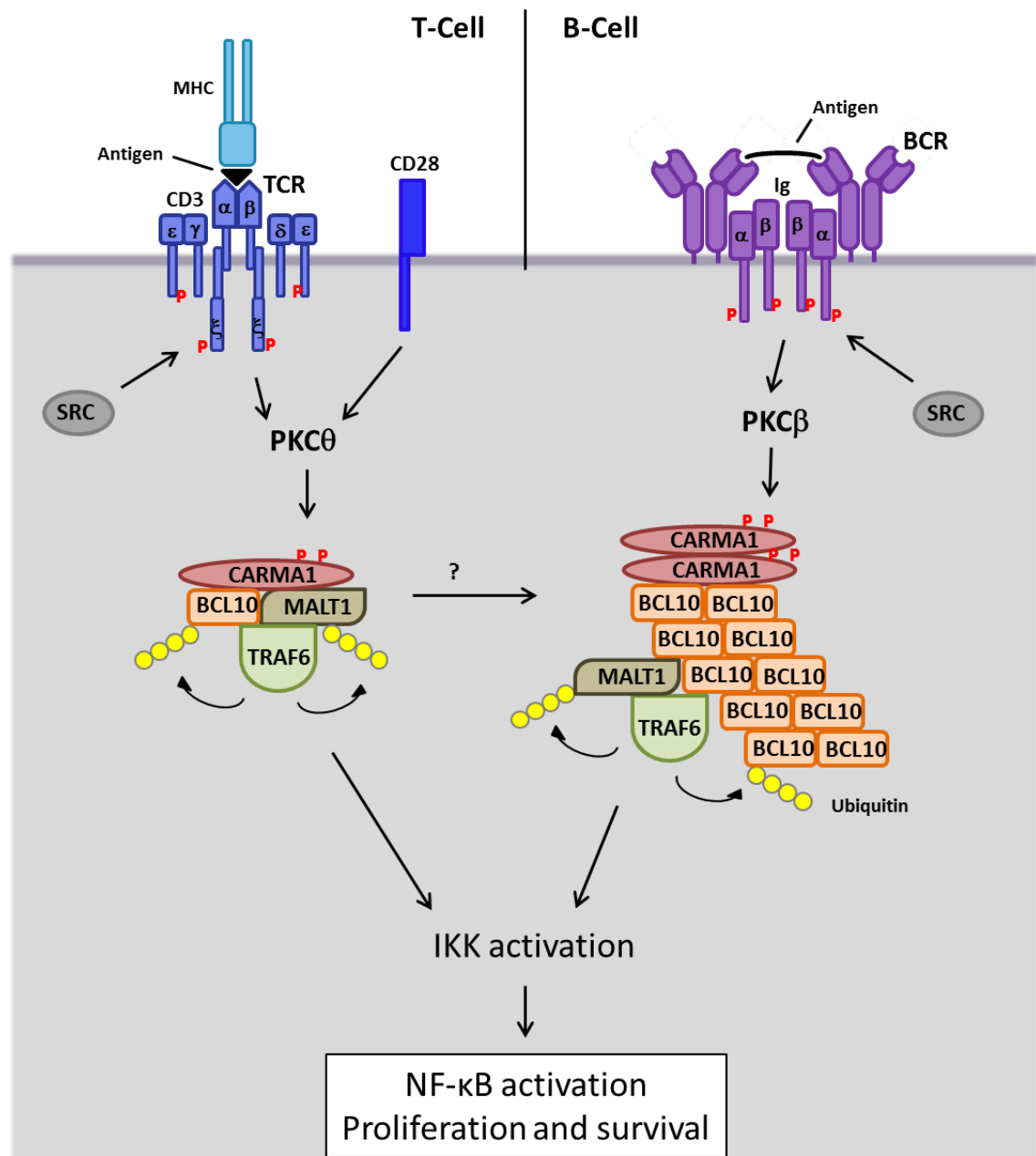


Figure 1-9 Simplified model of early signalling events in AgR mediated NF-κB activation. TCRs with co-receptor/stimulator CD28 and BCRs are triggering the activation of protein kinase C isoforms (PKCθ and PKCβ, respectively), which phosphorylate CARMA1 to enable recruitment of BCL10 via its CARD domain. It was recently suggested that in BCR signalling, CAMRA1 form clusters which nucleate the filamentous assembly of BCL10. If this applies to T-cells needs to be investigated. MALT1 interacts with the Serine-/Threonine-rich domain of BCL10 to connect the activated CARMA1 with the ubiquitin ligase TRAF6. Poly-ubiquitination events and TRAF6 are recruiting and activating the IKK complex. This leads to NF-κB translocation to the nucleus to induce transcription of genes involved in regulation of proliferation and cell survival. Figure based on Thome *et al.* (2010) and Qiao *et al.* (2013) [96, 103].

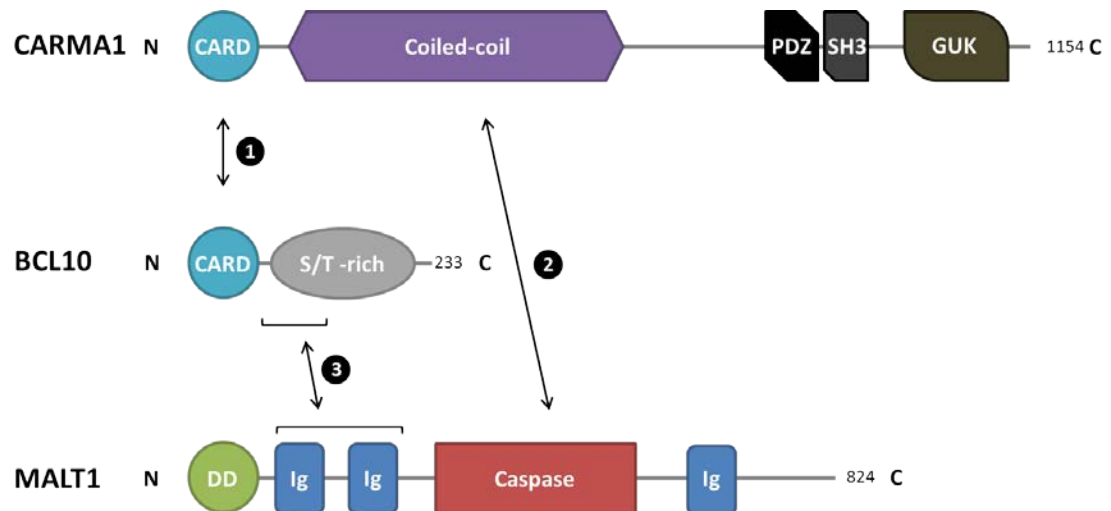


Figure 1-10 Cartoon of the domain structure of CARMA1, BCL10, and MALT1. Interaction 1: In 2001, Bertin *et al.* demonstrated the interaction between CARMA1-CARD and BCL10-CARD *via* co-immunoprecipitation and *in vitro* binding assays [97]. Interaction 2: Che *et al.* showed in 2004 with a two-hybrid approach that MALT1 directly binds to the coiled-coil domain of CARMA1 [104]. Interaction 3: In 2001, studies of Lucas *et al.* have shown that MALT1 binds to BCL10 through its immunoglobulin domains [25]. CARD: caspase recruitment domain, DD: death domain, S/T -rich: Serine/Threonine rich region, Ig: immunoglobulin-like domain, PDZ: post synaptic density protein (PSD95) - Drosophila disc large tumor suppressor (Dlg1) - zonula occludens-1 protein (zo-1), SH3: Sarcoma homology 3, GUK: guanylate kinase.

1.5.3.2 CBM-L complex formation and downstream signalling

After activation, CARMA1 recruits the adaptor protein BCL10 *via* CARD-CARD interaction. Similar to the CBM-M signalosome described in section 1.2.2.1, MALT1 is recruited to the CARMA1-BCL10 complex where the Ser/Thr-rich C-terminal part of BCL10 associates with the immunoglobulin-like domain of MALT1 [25]. Previous study has revealed that overexpression of BCL10 induces formation of cytoplasmic filaments shown by fluorescence microscopy in lymphomas, which are necessary for the recruitment of downstream effector molecules [105]. Qiao *et al.* (2014) recently

revealed the EM analysis of a helical filamentous BCL10 assembly. They also indicate that CARMA1 is forming clusters, which are nucleating BCL10 filament formation [103]. If this filament formation applies to TCRs as well, need to be investigated. So far, TCR signalling also was related to the formation of punctate structures named 'polkadots', which contain CBM molecules and appear to be critical for NF- κ B activation. [106].

Regarding downstream signalling of the CBM signalosome, MALT1 contains a binding site for the ubiquitin ligase TRAF6, which is required for IKK recruitment: The association of MALT1 with TRAF6 results in the addition of Lys63-linked (non-degradative) ubiquitin chains to a multitude of proteins, including BCL10 and MALT1, which can provide a docking surface for the recruitment of the IKK regulatory subunit NEMO. However, the mechanism for the subsequent activation of the IKK complex is still unclear [107]. It is possible that TRAF6 recruits the kinase TAK1, which activates IKK by phosphorylation [108]. IKK activation leads to NF- κ B activation, resulting in gene transcription to express proteins essential for proliferation and cell survival.

1.6 The death domain superfamily

In the above paragraphs, we have come across a variety of proteins controlling immunity, inflammation or apoptosis. This variation which includes receptors, adaptors, scaffold and effector proteins, often carry a motif belonging to the death domain superfamily. Members of death domain superfamily are responsible for the transmission of extracellular signals or from subcellular compartments to specific multiprotein platforms in signalling pathways of both, the innate (inflammasomes, apoptosomes, MyDDosomes *etc.*) and the adaptive (CBM complexes) immune system. They are key players in mediating protein-protein interactions important for transducing signals leading to apoptosis and inflammation, which has given them the this sinister name 'death domain'. Both, apoptosis and inflammation, are associated with many human diseases, which give studies of the death domains their ultimate biological importance. The members of this family can be categorised into four subfamilies: the death domains (DDs), the death effector domains (DEDs), the pyrin domains (PYDs) and the caspase recruitment domains (CARDs). These four types of domains are evolutionary conserved in organisms from mammals to *C. elegans* and *Drosophila*. Even viral proteins contain sequences that show high homology with the death domain superfamily, enabling them to interfere with host signalling pathways and suppress the apoptotic programme [109]. In humans, 102 family members have been discovered so far; 39 DDs, 8 DEDs, 33 CARDs and 22 PYD, which makes them one of the largest domain superfamilies [110, 111]. Proteins of the death domain superfamily contain a structurally conserved domain, comprising six amphipathic α -helices, arranged in an antiparallel α -helical bundle with Greek key topology, forming a hydrophobic core containing nearly all the conserved residues of this

superfamily [112-115]. However, there are small structural differences between the four subfamilies resulting in a change of the orientation of the helices as well as the electrostatic potential of their surface or the distribution of hydrophobic residues. Representative structures of the four subfamilies are shown in Figure 1-11. A structural comparison between the subfamilies show distinct characteristics including a more flexible helix 3 in DDs, whereas DEDs have conserved surface-exposed hydrophobic patches, CARs contain a kink in their first helix and PYDs contain a small helix 3 with an extended loop serving as linker between helix 2 and 3.

Studies have shown that members of the death domain family are difficult to work with because they have a tendency to aggregate. Only recently, we have gained more structural insights: Several death-fold structures have been solved, a number of them by using non-physiological conditions, solubility enhancement tags or by co-expressing them with another domain [110].

Interactions within the superfamily are almost all homotypic, which means interaction partners belong to the same subfamily. Compared to the number of structures available of isolated death domains, the number of complex structures solved is still very limited and it remains under debate if there is a common mode of interaction within the superfamily or the subfamilies. This debate has taken a significant turn over the time of this study, which will be discussed in the following sections.



Figure 1-11 Tertiary structures of the four death-fold subfamilies. The death domain superfamily comprises four subfamilies: Death domains (DD), death effector domains (DED), caspase recruiting domains (CARD) and pyrin domains (PYD). Highlighted are the structural characteristics of the individual subfamilies. Figure reprinted from Kersse *et al.* 2011 [9] with permission from Elsevier. PDB code for DD: 2OF5, DED: 2BBR, CARD: 3YGS, PYD: 1UCP.

1.6.1 Complex formation within death domain subfamilies

Back in 2001, only ten structures of the death domain superfamily with two complex structures were known: the 1:1 complex of APAF1-CARD and Procaspase-9-CARD [116] as well as the DD complex of the proteins Pelle and Tube. Weber and Vincenz (2001) [117] analysed and compared them, with the suggestion that the homotypic

interactions (DD-DD, DED-DED, CARD-CARD, PYD-PYD) of death-fold domains engage in three distinct types of asymmetric interactions to enable the formation of higher order complexes: A type 1 would be formed when residues from helices 1 and 4 of one death-fold domain interact with residues from helices 2 and 3 of another. This interaction was already represented by the CARD-CARD interaction between APAF1 and Procaspase-9 (Figure 1-12), where an acidic patch on APAF1 (helices $\alpha 2$ and $\alpha 3$) and a basic patch on Procaspase-9 (helices $\alpha 1$ and $\alpha 4$) are forming a 1:1 complex.

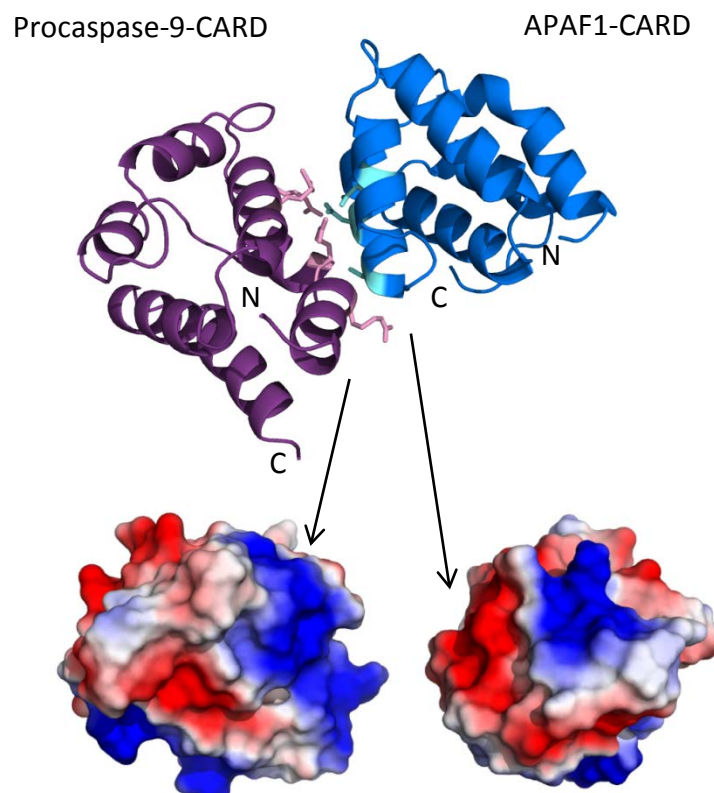


Figure 1-12 Structure of Procaspase9-CARD in complex with APAF1-CARD. Top: Crystal structure of the homotypic CARD-CARD interaction between Procaspase-9- and APAF1-CARD. Residues involved in complex formation are highlighted to demonstrate the interface between the two molecules. Bottom: Electrostatic representation of the binding interface. Procaspase-9 uses a positively charged surface (blue) to interact with a negatively charged surface (red) of APAF1-CARD. Prepared in PyMOL using PDB 3YGS.

The type 2 interaction was predicted to involve helix 4 and the loop between helices 4 and 5 of one death-fold domain interact with residues of the loop between helices 5 and 6. A type 3 interaction would be formed when residues from helix 3 of one death-fold domain interact with residues located on the loops between helices 1 and 2 and between helices 3 and 4 of the other death-fold domain. With these three interaction surfaces, a death-fold domain can potentially engage up to six binding partners. Since the prediction that the death domain superfamily may form these interactions, experimental data involving higher-order complexes and filamentous structures have confirmed that several subfamilies are capable of all three types of interactions, providing six interfaces per molecule. In the following sections, these structures will be presented in more detail.

1.6.2 The death domain (DD) subfamily

Members of the death domain subfamily earned their name for being involved in signalling pathways leading to caspase activation which ultimately results in apoptosis. However, nowadays it is well-known that DDs are just as important in other signalling cascades, such as those leading to NF- κ B activation during an inflammatory response. 15 out of 37 isolated DD structures have been solved up so far. Because of their tendency to aggregate, several structures had to be solved using extreme pH and/or by inserting de-aggregating mutations [118, 119]

1.6.2.1 Intermolecular interaction

Detailed molecular insight into how death-fold proteins mediate the formation of higher-order complexes has been first gained from crystal structures containing DD

proteins forming signalling platforms. One of these DD signalling platforms is the MyDDosome, a TLR induced assembly consisting of the adaptor molecule MyD88 as well as IRAK2 and IRAK4, that mediates the activation of NF- κ B. More details about the TLR signalling cascade can be found above in Figure 1-1 and section 1.2.1. In 2010, Lin *et al.* [17] published the crystal structure of the MyDDosome assembly at a resolution of 3.4 Å, which includes the death domains of six MyD88, four IRAK-4 and four IRAK-2 molecules (Figure 1-13).

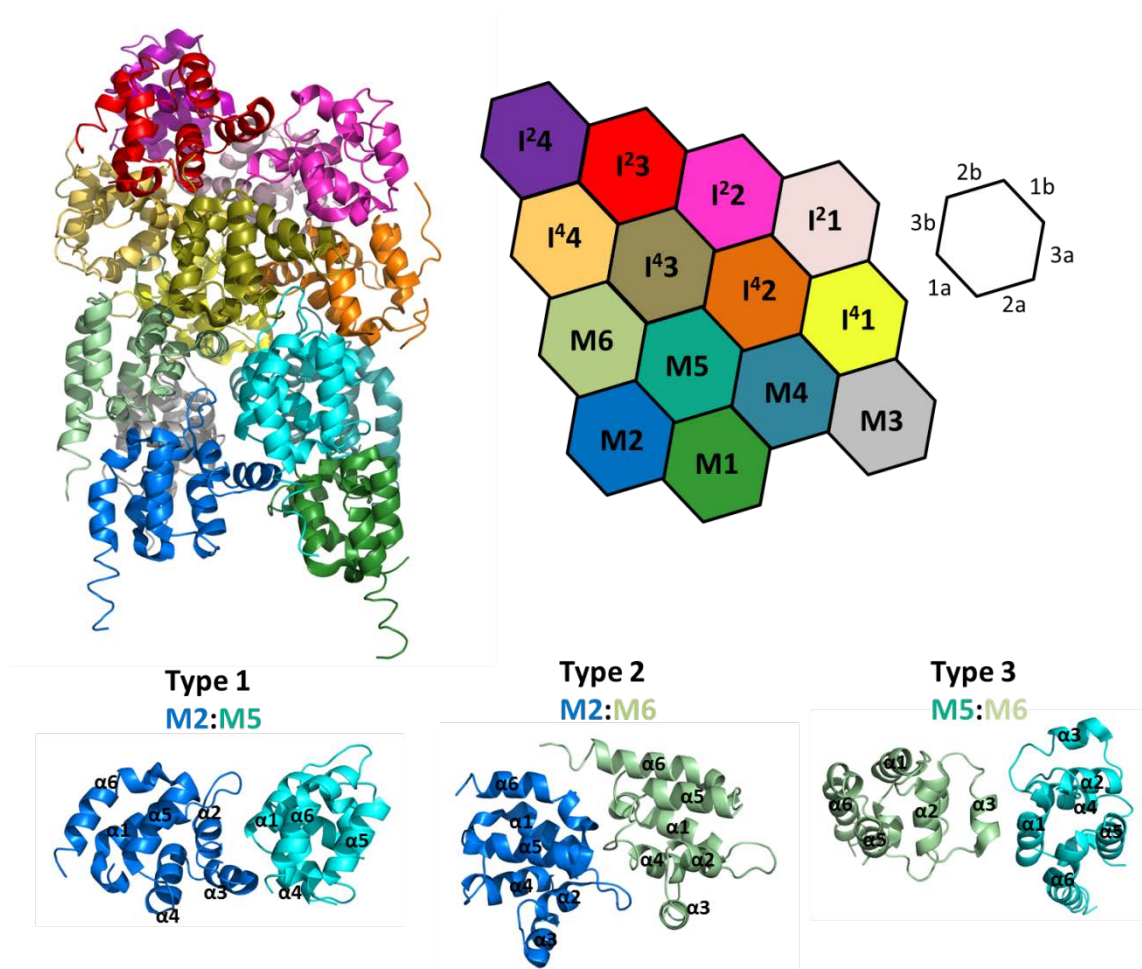


Figure 1-13 Ternary structure of the MyDDosome. Shown is a ribbon diagram of the MyDDosome, containing the molecules MyD88, IRAK2 and IRAK4. Schematic diagram on the right shows the three types of interactions in the complex (top). Given are the superimpositions of the three type 1 interactions, the three type 2 interactions and the five type 3 interactions in the ternary complex (bottom). Figure based on Lin *et al.* 2010 [17] and redrawn with PDB 3MOP in PyMOL.

The subunits of the complex are arranged in a left-handed helix with 3.7 subunits per turn, with type 1 and type 2 interactions mostly mediating the contacts between the layers, whereas type 3 interaction form contacts between adjacent DD molecules in the helical spiral. Based on the structure, it has been suggested that the assembly is nucleated by the MyD88 complex, that recruits the four IRAK4 molecules, followed by four IRAK-2 subunits. As the bottom surface of IRAK2 finds its best match in the top surface of IRAK4, IRAK2 is only recruited when IRAK4 is present on the top surfaces of MyD88. This is a cooperative process with every DD recruited increasing stability.

The MyDDosome is not the only higher-order complex formed by DDs. Another DD signalling platform is the PIDDosome, a p54-induced multiprotein complex formed upon DNA damage, comprising PIDD and RAIDD, which recruit and activate Caspase-2 resulting in apoptosis, DNA repair or NF- κ B activation [120]. Additionally, a further apoptosis inducing DD platform is called DISC (Death-Inducing Signalling Complex), which is formed upon activation of transmembrane death receptors such as FAS and TNFR (tumor necrosis factor receptor), which recruit the adaptor protein FADD (FAS-associated protein with death domain) via DD-DD interaction. FADD also contains a DED domain additionally to its DD. Activation leads to the recruitment of Procaspase-8 via DED-DED interactions, which completes the DISC assembly.

1.6.3 The death effector domain (DED) subfamily

In contrast to the death domains, no multiprotein assemblies of DEDs have been identified yet. On the other hand, whereas DDs are found as single copies in single- or multi-domain proteins, tandem DEDs exist in proteins such as Caspase-8 and -10, and viral Caspase-8/FLICE-inhibitory protein (vFLIP) MC159. The tandem DED containing

FLIPs are regulating DISC assembly and caspase activation. Similarly, MC159, a protein from poxvirus, is able to interact with FADD and Caspase-8 to inhibit apoptosis [121]. Currently, six DED structures are available. This includes the isolated domains of the above mentioned FADD-DED [113], the tandem DEDs of MC159 [121], PEA-15-DED [122] and the newest addition: the two Caspase-8-DEDs [123]. The structures of the isolated domains have revealed the specific characteristics of this subfamily: A conserved E/D-RxDL Φ (x=any amino acid, Φ =any hydrophobic amino acid) motif in helix 6 and the preceding loop (Figure 1-11 b) form a hydrogen-bonded charged triad with an acidic residue in helix 2. This feature is essential for DISC formation as it allows FADD self-association and MC159 binding. Additionally, another DED specific feature are conserved residues forming a hydrophobic patch on helix 2. This hydrophobic patch facilitates the FADD-Caspase-8 DED-DED interaction as well as the MC159 DED1-DED2 interaction [124].

1.6.3.1 MC159

The mechanism by which MC159 is inhibiting apoptosis is by interacting with FADD, therefore blocking FADD self-association, which prevents Caspase-8 to dock and induce downstream signalling. The DEDs of MC159 are locked in a compact, dumbbell-shaped structure, covering an extensive interface (Figure 1-14 A). This intramolecular interaction is mainly composed of hydrophobic interactions, involving helices $\alpha 2$ and $\alpha 5$ of DED1 and $\alpha 1$ and $\alpha 4$ of DED2. Therefore, this cannot be categorised as either a type 1, 2 or 3 interaction usually found in the death-fold superfamily. The reason for this difference is that DED1 is highly divergent from a typical DED fold; the topologically equivalent helix $\alpha 3$ is missing and is replaced by a

short loop connecting helices $\alpha 2$ and $\alpha 4$. Additionally, DED1 contains a short helix $\alpha 0$ at the N-terminus and helix $\alpha 7$, which helps to bury the hydrophobic core of the domain. The DED-DED intramolecular interaction involves the conserved hydrophobic patch on DED1, whereas the corresponding hydrophobic patch on DED2, as well as the charged triads on both DED1 and DED2, are free to interact with other DED-containing proteins [121, 124]

1.6.3.2 Caspase-8

The most recent crystal structure of a member of the death domain superfamily contains the tandem DEDs of Caspase-8 (Figure 1-14 B). Mutations F122A and I128D needed to be introduced to obtain soluble proteins. Similar to MC159, intramolecular interaction between DED1 and DED2 are mainly formed by hydrophobic residues from helices $\alpha 2$ and $\alpha 5$ of DED1 and the helices $\alpha 1$ and $\alpha 4$ of DED2. Whereas the overall structures of MC159 and caspase-8 are similar, there are variations which potentially define their individual function. An overlay of both tandem domains is provided and explained in Figure 1-14 C [123].

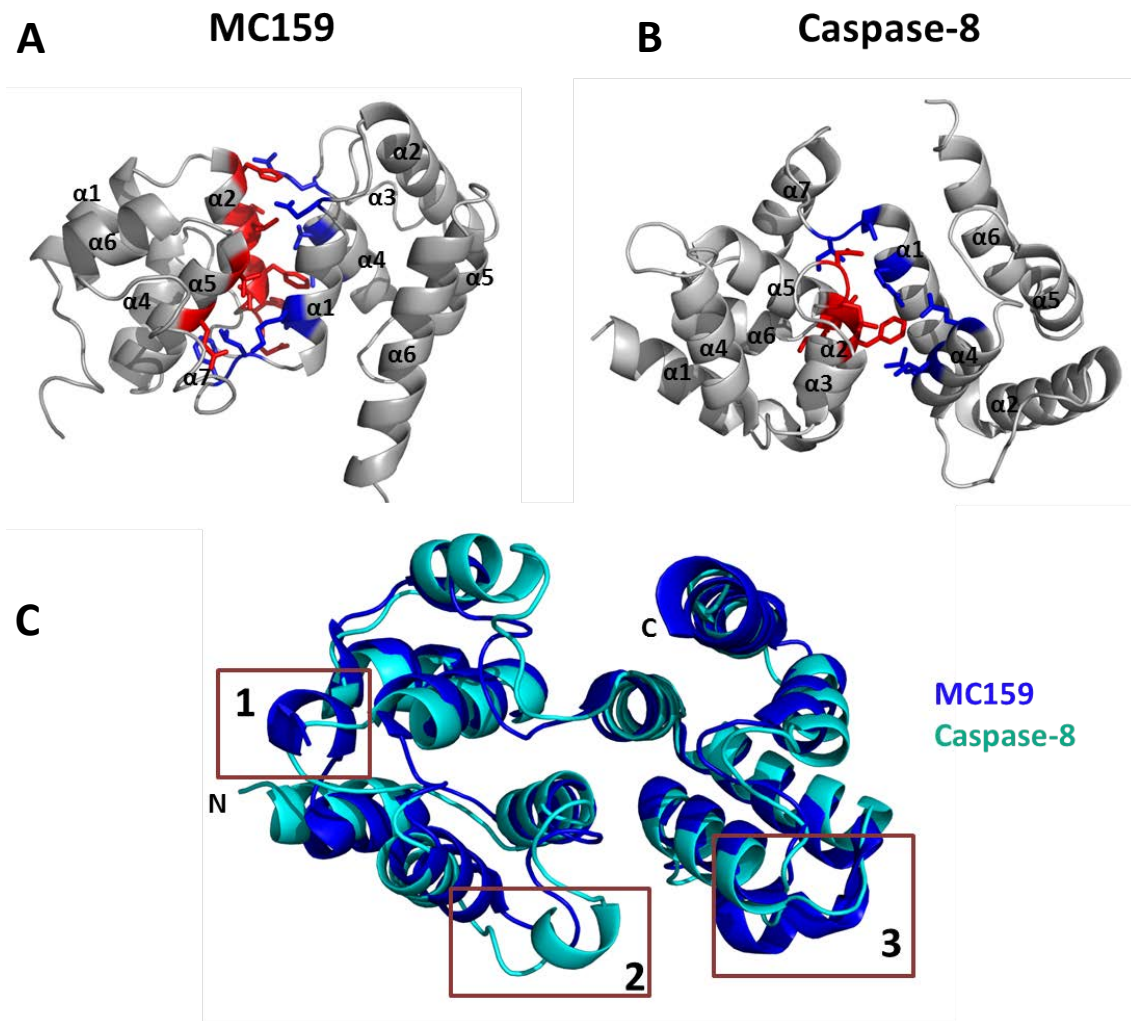


Figure 1-14 Structure of tandem death effector domains. A: Intramolecular interaction of the MC159 tandem DEDs. Helices $\alpha 2$ and $\alpha 5$ of DED1 interact with helices $\alpha 1$ and $\alpha 4$ of DED2. B: Tandem DEDs of Caspase-8 with helices $\alpha 2$ and $\alpha 5$ of DED1 and the helices $\alpha 1$ and $\alpha 4$ of DED2 forming an intramolecular interaction. C: Overlay of MC159-DEDs (blue) and Caspase-8-DEDs (cyan). Differences include such as helix $\alpha 0$ is missing in Caspase-8 (box 1) or a single turn helix in DED1 of Caspase-8 (box 2) whereas helix $\alpha 3$ in MC159 is replaced by a loop in DED2 of Caspase-8 (box 3). Figures were prepared in PyMOL using PDB code 2BBR for MC159 and 4ZBW for Caspase-8.

1.6.4 The pyrin domain (PYD) subfamily

The pyrin domains are the latest addition to the death-fold superfamily [125]. Their strongest characteristic feature is an altered helix $\alpha 3$. It can vary from being replaced by a long, exposed loop in NALP1 [114] to a short helix as seen in ASC [126]. So far, 9 of the 22 isolated pyrin structures have been solved. However, PYD-PYD interactions are not very well characterised and no complex structure is available at present. Helix $\alpha 3$ has been demonstrated to be essential in DD interactions, therefore $\alpha 3$ in pyrans may significantly alter the way PYDs undergo complex formation [117]. Proteins comprising a PYD as an effector domain define the NLRP (NALP) subclass of NLR proteins, which form inflammasomes. These proteins have been described in detail in section 1.3.2.

1.6.5 The caspase recruitment domain (CARD) subfamily

CARDs are playing important roles in innate and adaptive immunity and are involved in a wide range of interactions to form complexes, that are required for intracellular signal transduction. The 33 identified human CARDs are present in a large variety of proteins, including over 100 kDa scaffold and receptor proteins to 10 kDa regulator proteins containing just the single CARD (Table 1-1). As described in previous paragraphs, CARD containing proteins are involved in signalling cascades downstream of NLRs, RLRs as well as the TCR and BCR, often contributing to the formation of large oligomeric complexes that are important in regulating apoptosis as well as NF- κ B activation. The function of specific CARDs is summarised in Table 1-1. In the last 5 years, the interest in these domains has significantly increased resulting in the structures of many individual CARDs being solved by NMR or X-ray crystallography. A full list of available CARD structures is provided in Table 1-2. This list highlights that,

similar to the other subfamilies of the death domain superfamily, CARDS tend to aggregate and a number of structures have been solved using solubilising mutations, non-physiological conditions or fusion tags. Apart from the typical six-helical death-fold, CARDS are characterised by a kink in helix $\alpha 1$ (Figure 1-11). As CARDS are crucial for many different signalling pathways, it is important to know more about the molecular features and determinants of specificity of the interactions CARDS are involved in.

Subfamily	Protein	Protein	Protein	Protein
SHORCUT CARD	TRAF1	TRAF1 CARD	TRAF1 CARD	TRAF1 CARD
	TRAF2	TRAF2 CARD	TRAF2 CARD	TRAF2 CARD
	TRAF3	TRAF3 CARD	TRAF3 CARD	TRAF3 CARD
	TRAF4	TRAF4 CARD	TRAF4 CARD	TRAF4 CARD
LONGHAIR CARD	TRAF5	TRAF5 CARD	TRAF5 CARD	TRAF5 CARD
	TRAF6	TRAF6 CARD	TRAF6 CARD	TRAF6 CARD
	TRAF7	TRAF7 CARD	TRAF7 CARD	TRAF7 CARD
	TRAF8	TRAF8 CARD	TRAF8 CARD	TRAF8 CARD
SHORCUT CARD	TRAF9	TRAF9 CARD	TRAF9 CARD	TRAF9 CARD
	TRAF10	TRAF10 CARD	TRAF10 CARD	TRAF10 CARD
	TRAF11	TRAF11 CARD	TRAF11 CARD	TRAF11 CARD
	TRAF12	TRAF12 CARD	TRAF12 CARD	TRAF12 CARD
LONGHAIR CARD	TRAF13	TRAF13 CARD	TRAF13 CARD	TRAF13 CARD
	TRAF14	TRAF14 CARD	TRAF14 CARD	TRAF14 CARD
	TRAF15	TRAF15 CARD	TRAF15 CARD	TRAF15 CARD
	TRAF16	TRAF16 CARD	TRAF16 CARD	TRAF16 CARD
SHORCUT CARD	TRAF17	TRAF17 CARD	TRAF17 CARD	TRAF17 CARD
	TRAF18	TRAF18 CARD	TRAF18 CARD	TRAF18 CARD
	TRAF19	TRAF19 CARD	TRAF19 CARD	TRAF19 CARD
	TRAF20	TRAF20 CARD	TRAF20 CARD	TRAF20 CARD
LONGHAIR CARD	TRAF21	TRAF21 CARD	TRAF21 CARD	TRAF21 CARD
	TRAF22	TRAF22 CARD	TRAF22 CARD	TRAF22 CARD
	TRAF23	TRAF23 CARD	TRAF23 CARD	TRAF23 CARD
	TRAF24	TRAF24 CARD	TRAF24 CARD	TRAF24 CARD

Table 1-1 Classification of CARD proteins. Shorthair and longhair CARD proteins don't contain another domain, although can contain an additional sequence or motif. An example is BCL10, which has a Serine-/Threonine rich motif following its CARD. Proteins containing an additional > 500 amino acids long sequence without an obvious functional domain are categorised as 'longhair-CARD'. Table reprinted from Kao *et al.* 2015 [127] with permission from Springer Science and Business Media.

Protein/Complex	Method	Species	Residues	Purification and crystallization conditions	PDB ID	Year
ASC	NMR	human	full-length	pH 3.8 to increase solubility	2KN6	2009
ICEBERG	NMR	human	full-length	pH 3.8 to increase solubility	1DGN	2000
NLRP1	X-ray	human	1379-1462	crystallized at pH 7.5	3KAT	N/A
NOD1	X-ray	human	5-108	in complex with Ubiquitin, crystallized at pH 8.6	4JQW	2014
			16-108	crystallized at pH 4.7	2NSN	2007
			14-106	crystallized at pH 6	2NZ7	2008
	NMR	human	15-138	residues 107-138 added to increase stability	2B1W	2007
RAIDD	NMR	human	1-136	residues 118-136 added to increase solubility	3CRD	1999
Apoptosome	X-ray	human	APAF1: 1-97	domains individually expressed in <i>E. coli</i> and mixed.	3YGS	1999
			Caspase-9: 1-112	crystallized at pH 4.6.		
		<i>c.elegans</i>	full-length CED4	domains individually expressed in <i>E. coli</i> and mixed.	3LQQ and 3LQR	2010
			full-length CED3	crystallization-pH was not indicated.		
		human	APAF1: 1-97 Caspase-9: 1-108	no apparent electron density for CED3 in the crystals. crystallized at pH 5.5	4RHW	2014
APAF1	X-ray	human	1-97	crystallized at pH 8	1CY5	1999
			1-97	crystallized at pH 4.6	2YGS	1999
			1-97	crystallized at pH 6.1	2P1H	2007
			full-length	crystallized at pH 7.1	1Z6T	2005
	NMR	human	1-97	pH 6.5	1CWW	1999
			1-97		1C15	1999
CED-4 (complexed to CED-9)	X-ray	<i>c.elegans</i>	full-length	co-expressed and co-purified	2A5Y	2005
clAP1	NMR	human	435-562		2L9M/3T6P	2011
CARD8	X-ray	human	full-length	crystallized at pH 5.2	4IKM	2013
BINCARD	X-ray	human	1-99	crystallized at pH 7	4DWN	2013
ARC	X-ray	human	1-95	crystallized at pH 8.2	4UZ0	2015
CARMA1	X-ray	human	21-108	mutation: E24D, dimer, at pH 6.5	4JUP	2013
			18-110	mutations: E18A, E20A, at pH 6.5	4I16	2012
			14-110	crystallized at pH 6.5	4LWD	2013
BCL10	NMR	human	1-115	mutation: E53R, at pH 6.5	2MB9	2013
RIG-I	NMR	human	90-190	pH 7.5	2LWD	2012
			90-190	mutation: T170E, at pH 7.5	2LWE	2012
	X-ray	duck	full-length	crystallized at pH 8.5	4A2W	2011
		human	1-188	tetramer with ubiquitin, pH 7	4NQK	2014
MAVS (IPS-1)	X-ray	human	3-93	N-terminally fused to MBP to increase solubility	2VGQ	2008
		horse	1-94	mutation: E26R	4O9L	2014
			1-94	mutation: R64C	4O9F	2014
MAVS filament	EM	human	1-93	9.6 Å - cryo EM (pH 7.5)	3J6C	2014
			1-97	3.6 Å - cryo EM (pH 7.5)	3J6J	2014
MAVS/RIG-I	X-ray	humn	RIG-I: 1-188 MAVS: 1-97	pH 7.5	4P4H	2014

Table 1-2 CARD structures available as up to October 2015. Information collected from <http://www.rcsb.org/pdb/>.

1.6.5.1 Tandem CARDs and their intramolecular interactions

CARDs and death effector domains are the only subfamilies with protein members containing tandem domains. Tandem CARDs are found in three different proteins: RIG-I [128], MDA5 [129] and NOD2 [64].

RIG-I and MDA5

RIG-I and MDA-5 are both members of the RLR family of PRRs and viral RNA sensor proteins. More information on their function can be found in section 1.3.1. The crystal structure of full length duck RIG-I (which shares 53% sequence identity with human RIG-I) has been solved in 2011 by Kowalinski *et al.* (Figure 1-15, cyan) [33]. This was followed by the crystal structure of the human tandem CARDs of RIG-I (overlaid in Figure 1-15, grey) in 2014 by Peisley *et al.*[37]. In these complex structures, the loop between helices 5 and 6 in CARDa and the loop between helices 4 and 5 of CARDb form an intramolecular interaction, which corresponds to a type 2 interaction. These two CARDs form a rigid unit and adopt a head-to-tail conformation, where the C-terminus of CARDa and the N-terminal helix of CARDb are linked together to form one long single helix. The interaction is based on both hydrophobic and polar interactions with Y24 and Y81 in CARDa and I130, S160 and K163 in CARDb as the main contributors. This structure underlines the finding that both CARDs serve as one functional unit and are both required for downstream signalling shown by Zang *et al.* in 2010 using IRF3 dimerisation as a read out.

So far, there is no crystal structure available for the tandem CARDs of MDA5. However, SAXS data shows a compact structure and suggests that they might interact with each other as seen in RIG-I [42].

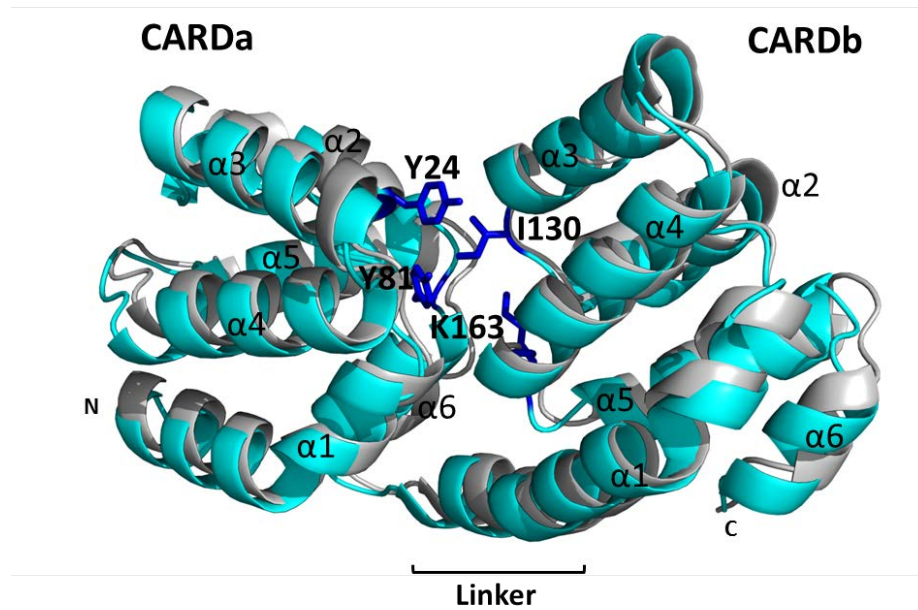


Figure 1-15 Structure of RIG-I tandem CARDs. Duck RIG-I is showed in cyan with the main residues of the interface highlighted in blue. As comparison, human RIG-I is overlaid and shown in grey (root mean square deviation: 1.0). PDB 4A2Q for duck and 4NQK for human RIG-I were processed in PyMOL.

NOD2

The third protein containing tandem CARDs is NOD2. Wagner *et al.* [130] first demonstrated that there is an interaction between the two NOD2 CARDs using yeast-two hybrid assays. Previous work in our lab has confirmed these results using purified proteins: ITC and pull-down studies have been performed to characterise the binding stoichiometry and thermodynamic properties of this interaction. ITC studies by titrating CARDa into CARDb resulted in a K_D of $\sim 1 \mu\text{M}$ at 25°C with a 1:1 stoichiometry. In addition, thermal unfolding experiments monitored by CD spectroscopy indicate that the two CARDs stabilise each other, which is shown in Figure 1-16. However, despite these studies, the exact binding surface involved in the intermolecular

interaction could not be identified and further studies are necessary to characterise the interaction.

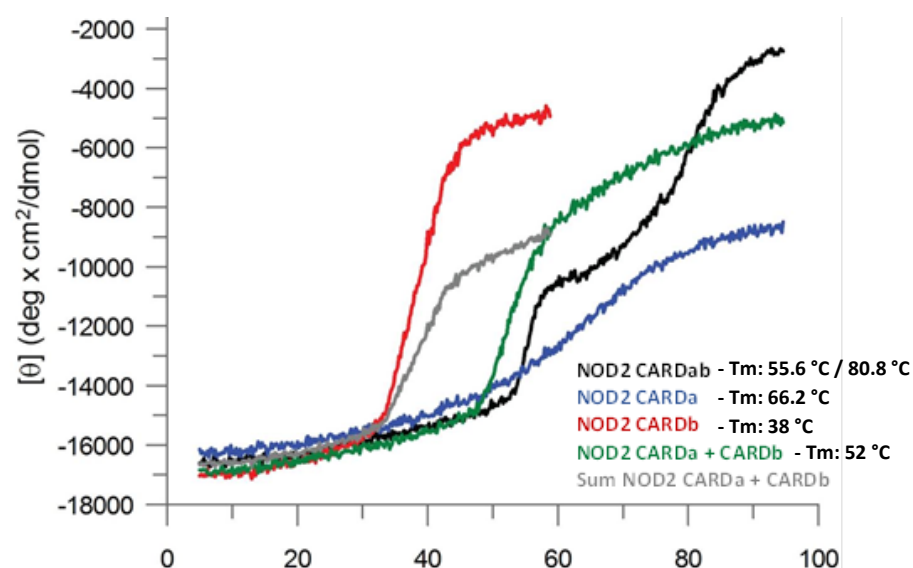


Figure 1-16 CD analysis of the NOD2 CARDa-CARDb complex. Thermal unfolding of the complex revealed a stabilising effect between CARDa and CARDb of NOD2. The complex (green) unfolded at a higher temperature (than the computed mean value (grey) of the individual curves. Figure from Fridh et al.(2012) [131].

NOD2 - preliminary work

The tandem CARDS of human NOD2 are well expressed and soluble. However, extensive crystallisation trials of the NOD2 tandem CARDS have been performed but did not result in any diffracting crystals. Therefore, the focus was set on the individual CARDS. In contrast to the tandem construct, expression and purification of each individual CARD has not been possible in quantities required for biochemical studies. Consequently, a specific protocol allowing the purification of the individual CARDS has been developed in the group [131]. A Glutathione S-transferase (GST)-fusion construct (pGEX-6P1; GE Healthcare) containing the native sequence of NOD2 aa 28-218 was

modified to contain a thrombin cleavage site (VPRGS) between residues leucine 119 and histidine 120. Following affinity chromatography, the GST-fusion protein was cleaved on-column with thrombin (Merck Millipore) to elute CARDb. With GST-CARDa and a large portion of cleaved CARDb remaining on the column, another on-column cleavage was performed using PreScission protease (GE Healthcare) to yield CARDa. An anion-exchange step (Hi Trap Q FF; GE Healthcare) was conducted with the eluted CARDa/CARDb mixture to eliminate CARDb contamination prior to Gel filtration (Superdex 75 16/60; GE Healthcare).

As the extensive crystallisation trials of the NOD2 CARDS (tandem and individual) have not resulted in any diffracting crystals, it was decided to use NMR spectroscopy to gain structural insights. While the tandem CARD construct was not amenable to NMR studies due to a tendency to dimerise at higher concentrations, the structure of NOD2-CARDa could be determined by Dr Kovilen Sawmynaden, shortly before leaving the group (Figure 1-17). The calculated ensemble of 10 lowest energy and water-refined structures show good convergence with an all-atom root mean square deviation (RMSD) of 0.34 Å. CARDa is comprised of five alpha-helices, which come together to form the familiar alpha-helical bundle. However, unlike all other structures of isolated CARDS determined to date, the sixth alpha helix is absent.

In order to investigate the interaction of CARDa with CARDb, NMR saturation transfer difference experiments were completed. This experiment is a popular tool, which allows the identification of ligand regions/residues important for binding. This method is based on the nuclear Overhauser effect between bound ligand protons and all protein protons. This experiment involves the comparison of protein-ligand complex spectra measured under on-resonance (I_{on}) and off-resonance (I_{off}) irradiation [132].

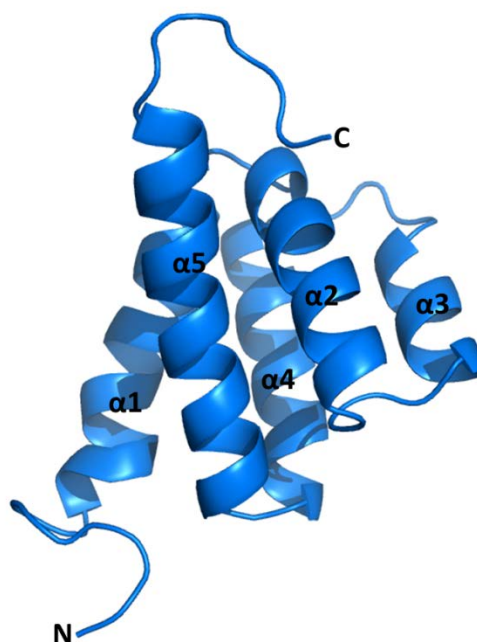


Figure 1-17 Solution structure of NOD2 CARDa. An average is shown of the calculated ensemble of 10 lowest energy and water-refined structures of CARDa (unpublished data; Rittinger Group).

In Figure 1-18, the off-resonance spectrum (light blue) is shown as the reference spectrum and superimposed with the difference $\Delta I = I_{\text{off}} - I_{\text{on}}$ (dark blue). Using this method, two tryptophan residues have been identified to be involved in CARDb binding. Following this result, titration experiments were performed, with the focus at the tryptophan region on the superimposed ^{15}N -HSQC spectra of CARDa alone and CARDa bound to CARDb (Figure 1-19). Full spectra can be found in Appendix A 1-1. This experiment suggested that W59, W63 or W68 are the most likely tryptophans to be located at the binding site. These experiments were the basis of this part of my project, which was aimed at investigating the potential role of these tryptophan residues in CARDb binding and further characterisation of the CARDa-CARDb interface.

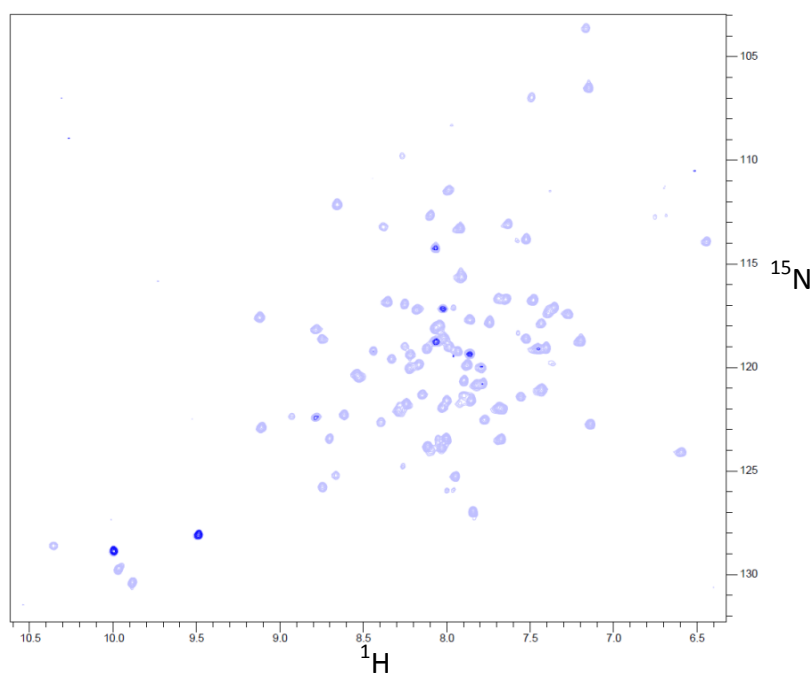


Figure 1-18 NMR Saturation transfer difference spectra. This experiment shows a sample of CARDa bound to CARDb. The reference spectrum (I_{off}) is shown in light blue and is superimposed with the difference $\Delta I = I_{\text{off}} - I_{\text{on}}$ presented in dark blue.

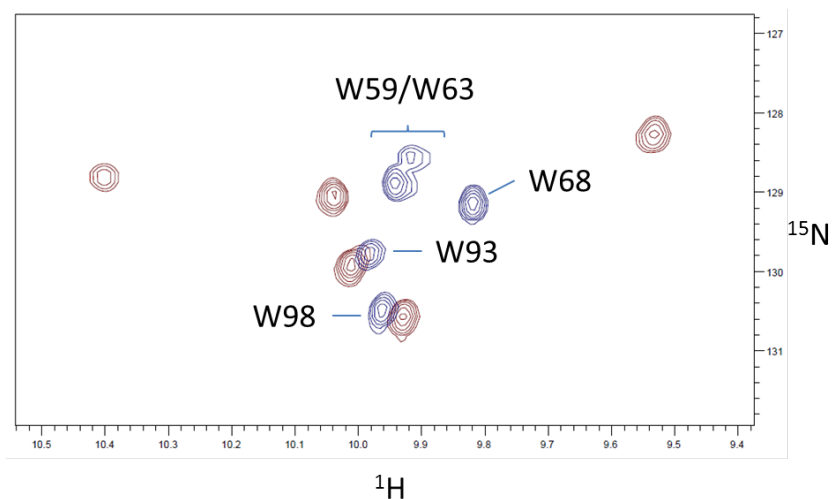


Figure 1-19 Tryptophans of NOD2. A section of the ^{15}N - ^1H TROSY-HSQC spectra of CARDa (red) and CARDa+CARDb (blue) show clear chemical shift changes in the tryptophan region (unpublished data; Ritinger Group).

1.6.5.2 Intermolecular interactions

In 1999, when the structure of the 1:1 complex of APAF1-CARD and Procaspase-9-CARD was published (Figure 1-12), probably nobody expected that this will remain the only real structural insight into intermolecular CARD-CARD interactions for the next 14 years. The discovery of the multi-protein death domain signalling platforms such as the PIDDosome or MyDDosome established the new idea that the formation of higher-order assemblies could apply to all the death-fold subfamilies. This idea was supported by the a cryo EM structure of the apoptosome in 2010 by Yuan *et al.* [54] (Figure 1-6) although the CARD-CARD interactions are not defined in this structure. However, apart from the 1:1 interaction and the higher-order assemblies, a third mode of interaction has recently become part of the discussed models: the formation of filaments.

CBM-L complex

Considering the structural similarities of CARDS and DDs, the idea that CARDS may also form multimeric protein complexes, seemed only logical. It was in 2013, when finally a structure showed type 1 and type 2 interactions involving CARDS: Qiao *et al.* [103] demonstrated the filament formation of BCL10 CARDS downstream of BCR in the NF- κ B signalling pathway via the CBM-L complex (Figure 1-20, more information about this pathway can be found in section 1.5.3). The solution structure of BCL10 (E53R) was solved before the BCL10 subunits were fitted into an EM density map at a ~ 20 Å resolution, which resulted in a modelled structure of a CARD filament (Figure 1-20). Via EM analysis of streptavidin gold-labelled biotinylated CARMA1 together with the

BCL10 filament, CARMA1 was determined to be only located at the tip of the filament, and was therefore suggested to nucleate the CARMA1/Bcl10 filament.



Figure 1-20 EM model of a BCL10 filament. A: Shown is the symmetry of the BCL10 filament forming a left-handed “four-start” helical strand. B-D: A closer look is given on the involved residues in the type 1 and type 2 interactions. E: EM density map is shown with the fitted BCL10-CARDs in cyan. Figure reprinted from Qiao *et al.* (2013) [103] with permission from Elsevier.

MAVS signalosome

The BCL10 filament structure was soon followed by other filamentous assemblies: the MAVS-CARD filament. EM analysis of MAVS-CARD by Hu *et al.* (2014) [133] resulted in a model at 3.6 Å resolution of a single-stranded left-handed helical arrangement. More importantly, the first crystal structure of a multi-domain CARD complex was revealed

by the same group. MAVS filament formation was limited by inserting mutations in either RIG-I and MAVS CARDS (Figure 1-21). This complex includes four RIG-I tandem CARDS, which are tightly joined together through the type 2 interaction, whereas type 1 and type 3 interactions were observed between the adjacent RIG-I units. Additionally a MAVS-CARD tetramer is stacked on top of the RIG-I tetramer through type 2 interactions. To underline these findings, this RIG-I/MAVS complex is arranged in the same way as the individually solved EM MAVS filament, which can be superimposed well with the RIG-I tetramer proposing a model of RIG-I nucleating MAVS-CARD filament formation.

NODosome

The tandem CARDS of NOD2 and their binding partner RIP2-CARD have previously been co-expressed in *E.coli* in order to carry out binding studies by pull-down experiments. These studies provided evidence for a direct interaction between NOD2 and RIP2. The two residues R38 and R86 in CARDa of NOD2 have been identified to play a crucial role in RIP2-CARD binding. The exact binding surface has yet to be determined and due to only a limited amount of mutants tested so far, it is still unclear whether CARDb participates directly, indirectly or not at all in the binding of RIP2. Further work will be explained in Chapter 4.

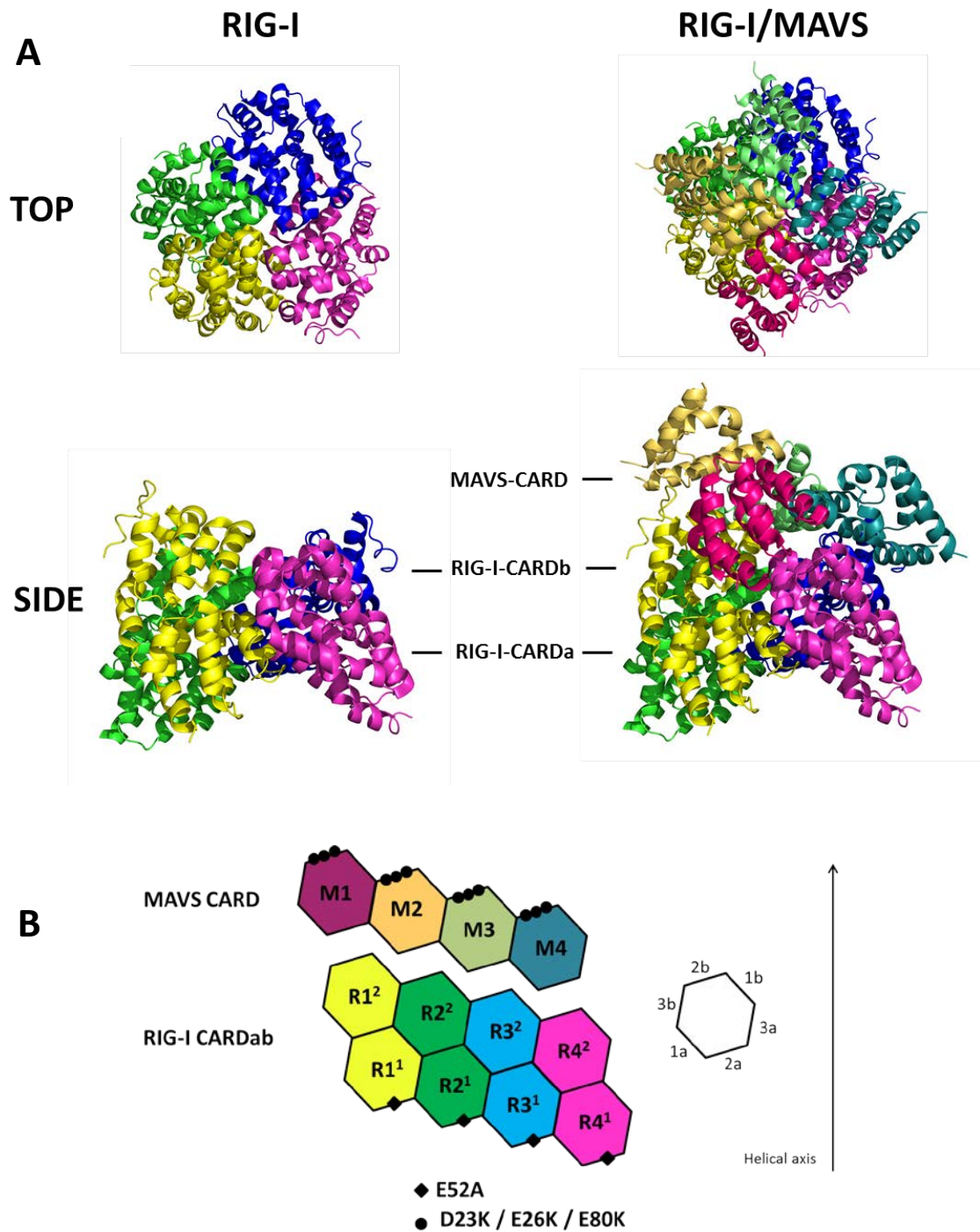


Figure 1-21 RIG-I-CARDb/MAVS-CARD complex. A: Mutational engineering allowed the crystallisation of a RIG-I-CARDb tetramer with four MAVS-CARD molecules stacked against the CARDb domains of RIG-I. B: 2D representation of the RIG-I-CARDb/MAVS-CARD complex represented with the mutations. Type 1 and type 3 interactions were observed between the adjacent tandem RIG-I units, whereas type 2 interactions were formed between MAVS and RIG-I CARDS. This is also highlighted in the schematic diagram on the right. Figure based on Wu *et al.* (2014) [133] with PDB 4P4H processed in PyMOL.

1.7 Project aims

CARDs are important protein interaction domains of the innate and the adaptive immune system. However, our understanding of the molecular details underlying CARD-mediated protein-protein complex formation is limited and we do not understand if there is a general mechanism by which CARDs interact and if their mechanism of complex formation is similar to other members of the death domain superfamily.

The overall aim of this thesis was to investigate the biochemical and biophysical characteristics of CARD-CARD interactions, to gain insights into the binding specificity and/or stoichiometry of CARD-mediated protein complexes and to elucidate if there is a common mode of interaction within this protein subfamily. The aim was to identify interacting CARD pairs that can be expressed and purified to then further investigate their binding characteristics. Specifically, we want to determine if CARDs form 1:1 protein complexes, or if CARDs are able to interact with up to six partners to form a signalling platform as seen with other members of the death domain subfamily, or if an additional oligomerisation domain is required to bring CARDs into close proximity to be able to interact in a similar manner as observed in the apoptosome. The ultimate goal was to gain structural insight by either crystallography or NMR spectroscopy but also to investigate biochemical and biophysical binding properties with available techniques including NMR spectroscopy, ITC, CD and fluorescence spectroscopy.

Chapter 2

Materials and methods

2.1 Materials

Chemicals and reagents used in this study were purchased from Agilent Technologies, Cambridge Isotope Laboratories, Expedeon, GE Healthcare, Generon, Merck, New England BioLabs (NEB), Novagen, Qiagen, Roche, Sartorius, Sigma Aldrich, Source Bioscience and Thermo Fisher Scientific (Thermo Scientific). Specific chemicals and reagents are detailed throughout this chapter, in their relevant sections. Oligonucleotide primers were obtained from either Eurogentec or MWG.

2.2 Molecular biology

2.2.1 Bioinformatics

Nucleotide sequences of target genes were downloaded from the European Nucleotide Archive website (<http://www.ebi.ac.uk/ena>). Protein domain boundaries were designed based on domain prediction and secondary structure prediction using the following websites: SMART (<http://smart.embl-heidelberg.de>), ScanProsite (www.expasy.ch/tools/scanprosite), JPred (www.compbio.dundee.ac.uk/wwwjpred/) and Uniprot (<http://www.uniprot.org/>). The Double Digest Finder from the New England BioLabs (NEB) website (<http://www.neb.com/nebecomm/DoubleDigestCalculator.asp>) was used to select correct buffers and conditions for DNA (deoxyribonucleic acid) restriction digests. The LIC primers were designed with the aid of the Protein Crystallization Construct Designer (Protein CCD) tool (<http://xtal.nki.nl/ccd/Welcome.html>). Mutagenesis primers have been designed using the QuikChange Primer Design (<http://www.genomics.agilent.com/primerDesignProgram.jsp>). Emboss needle (http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html) was used to verify results of cloning/mutagenesis

experiments by aligning sequenced DNA with template DNA. Biophysical parameters of cloned proteins such as isoelectric point (pI), extinction coefficient (ϵ) and molecular weight (M_w) were determined using the PROTParm calculator on the Expasy server (www.expasy.ch). Protein sequence alignments were generated using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). Protein sequence alignments were visualised and edited and processed using JalView (<http://www.jalview.org/>) Pairwise protein sequence identities/similarities were calculated using Emboss align (<http://www.ebi.ac.uk/tools/emboss/align/>). SWISS-MODEL server (<http://swissmodel.expasy.org/>) was used for protein structure homology modelling.

2.2.2 Construction of expression plasmids

Details of cloned and purified proteins used in this study can be found in the relevant results sections. Plasmids for *E.coli* expression of recombinant proteins were generated by PCR amplification of the target gene followed by ligation independent cloning (LIC) or infusion cloning. Human cDNAs of CARD-containing proteins served as templates for polymerase chain reaction (PCR) and were obtained from Pascal Meier (ICR, UK), Margot Thome (UNIL, CH) or Life Technologies (<https://www.lifetechnologies.com>). Primer design was carried out taking into account the following parameters: 40-60 % GC content, ~30 bases in length and melting temperature similar for both forward and reverse primer between 55-72 °C ($T_M = 2 \times (A+T) + 4 \times (G+C)$). The nucleotide sequence 'CAGGGACCCGGT' (forward) and 'GGCACCAGAGCGTTA' (reverse) or 'CTTTCAGGGACCCGGT' (forward) and 'CAGAATTCGGATCCTGTTA' (reverse) were added to the primers for LIC cloning or infusion cloning, respectively. Primers for site-directed mutagenesis were designed

using the Agilent Technologies' QuikChange® Primer Design Program (see section 2.2.1 for website). Primers were ordered from either Eurofins or Sigma Aldrich. The DNA fragments have been cloned in three different LIC vectors: pET52-47, designed by Vangelis Christodoulou (The Francis Crick Institute), which contains an ampicillin resistance and generates N-terminally His₆-tagged proteins. pET49b (Novagen), a vector containing kanamycin resistance that generates His₆-GST-fusion proteins. pGST (Vangelis Christodoulou), which contains a kanamycin resistance and generates GST-fusion proteins. For infusion cloning, pHMBP (Vangelis Christodoulou) was used to generate a N-terminally His₆-tagged MBP (maltose binding protein) fusion proteins. All vectors include a 3C-protease cleavage site for tag removal. Plasmid vector maps can be found in Appendix A 2-1.

2.2.3 Polymerase chain reaction (PCR).

Template DNA was amplified using an Eppendorf Mastercycler® DNA Engine Thermal Cycler. The reaction was performed using the KOD Hot Start DNA polymerase kit (Merck) with a 50 µL reaction volume (Table 2-1). The PCR protocol consisted of 40 cycles of strand separation, primer hybridisation and DNA synthesis (Table 2-2). The PCR purification was performed using QIAquick PCR purification kit (Qiagen) as per manufacturer's instructions. DNA was eluted using 50 µL of distilled water.

Reagent	Volume (µL)
2x KOD Hot Start master mix	25
dsDNA template (20ng/µL)	1
Forward primer (10uM)	1
Reverse Primer (10uM)	1
ddH ₂ O	22
Total Volume	50

Table 2-1 Typical PCR reaction mixture for cloning.

Step	Temperature (°C)	Time (sec)
1. Polymerase activation	95	120
2. Denaturation	95	20
3. Annealing	55	10
4. Extension	70	20/kb
Repeat steps 2-4 39 times		
5. Final extension	70	120
6. Cooling	10	300

Table 2-2 Typical PCR cycling protocol for cloning.

2.2.4 Agarose gel electrophoresis

DNA fragment size of the PCR product was determined by agarose gel electrophoresis and compared with known DNA size markers such as 1 kb or 100 bp DNA Ladders (NEB). Gels were prepared by dissolving 0.8% (w/v) agarose (Thermo Scientific) in 1X TAE buffer (40 mM Tris Acetate pH 7.7, 1 mM EDTA), whereas SYBR® Safe DNA Gel Stain (Thermo Scientific) was added to the gels to allow visualization of DNA bands by ultraviolet (UV) light at 254 nm.

2.2.5 Determination of DNA concentration

The NanoDrop® ND-1000 spectrophotometer (Thermo Scientific) was used to determine Plasmid DNA concentrations by UV spectrophotometry with $1 A_{260} = 50 \mu\text{g/mL}$. Alternatively, the intensity of visualized DNA bands was used to estimate concentration by comparison with DNA markers of known concentration.

2.2.6 Restriction enzyme digestion

In preparation for ligation with the desired DNA fragment, LIC vectors have been digested using the compatible restriction enzymes KpnI and SacI (NEB), which was performed at 37 °C for 2 hours (Table 2-3). On the other hand, the vectors for infusion cloning were digested using only KpnI. Following restriction digestion, the 20 µL reaction sample was mixed with 5 µL 5x Loading Buffer (Qiagen) and run on a 0.8 % agarose gel. The DNA band was cut out from the gel and extracted using the QIAquick Gel Extraction kit (Qiagen) following manufacturer's instructions and eluted with 35 µL dH₂O.

Reagent	Volume (µL)
Purified plasmid	10
10X BSA (if required)	2
10X reaction buffer	2
Restriction enzyme 1 (10U/µL)	1
Restriction enzyme 2 (10U/µL)	1
ddH ₂ O	4
Total volume	20

Table 2-3 Typical digestion protocol.

2.2.7 Ligation independent cloning

An overview is provided in Figure 2-1. The purified PCR products containing the LIC sites were treated with T4 DNA polymerase (Thermo Scientific) adding dATP to produce sticky overhangs (Table 2-4). Reaction as carried out at 37 °C for 30 min, followed by heat inactivation at 70 °C for 20 min. Cut and purified vectors were treated similarly, although the T4 treatment was carried out in the presence of dTTP instead of

dATP (Thermo Scientific), followed by gel extraction purification. LIC ready DNA fragment and vector were mixed together (3:1, [insert]:[vector]) and incubated for 5 minutes at room temperature, followed by adding 1 μ L of EDTA. An overview of the single steps is provided in Figure 2.1.

Reagent	Volume (μ L)
Purified DNA fragment	3
dATP	2
NEB Buffer 2.1	2
ddH ₂ O	12
T4 polymerase	1
Total volume	20

Table 2-4 Typical T4 treatment protocol.

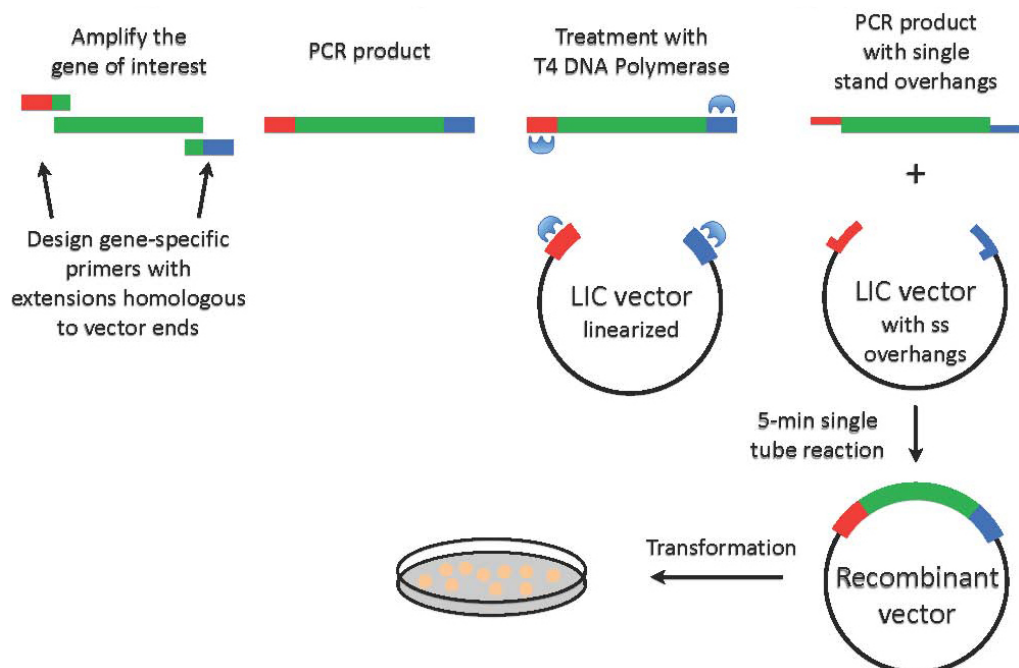


Figure 2-1 Ligation-independent cloning steps. Diagram illustrating the steps involved in ligation-independent cloning. Figure adapted from Aslanidis (1990) [134].

2.2.8 Infusion cloning

Infusion cloning was carried out using the In-fusion®HD cloning kit (Clontech® Laboratories) (Table 2-5). Purified PCR products containing the infusion cloning sites do not need any further treatment. Vector and insert quantity was calculated using the Clontech Laboratories' Molar Ratio Calculator (<http://bioinfo.clontech.com/infusion/molarRatio.do>). The reaction was carried out at 50 °C for 15 min.

Reagent	Volume (µL)
5X In-Fusion HD Enzyme Premix	2
Linearized Vector (200ng/µL)	1
Purified DNA fragment (50ng/µL)	1
ddH ₂ O	6
Total volume	10

Table 2-5 Typical Infusion cloning reaction.

2.2.9 Transformation of cloning products

LIC and Infusion products were transformed into competent NovaBlue GigaSingles™ *E.coli* cells (Novagen) by heat-shock treatment at 42 °C for 30 seconds followed by an incubation time on ice for 60 seconds. After incubating the cells at 37 °C for an hour, they were plated on LB-agar, containing the appropriate antibiotic resistance (50 µg/mL kanamycin, 50 µg/mL carbenicillin and/or 34 µg/mL chloramphenicol), followed by a further incubation over night at 37 °C (Table 2-6).

Step	Protocol
1	Thaw competent cells on ice (50 μ L/tube)
2	Add 3 μ L LIC / infusion reaction mix to each tube
3	Incubate on ice for 30 minutes
4	Heat shock cells for 45 seconds at 42 °C
5	Incubate on ice for 60 seconds
6	Add 500 μ L SOC to each tube.
7	Incubate at 37 °C for 1 hour
8	Plate 250 μ L on LB-agar plates supplemented with appropriate antibiotics
9	Incubate plates at 37°C over night
10	Pick 2-5 single colonies for colony PCR

Table 2-6 Transformation protocol using heat shock.

2.2.10 Identification of positive plasmids and purification

Colony PCR using the Kapa2G PCR ready mix kit (Kapa Biosystems) was performed to screen for plasmids containing the desired insert. The reaction mixture of 25 μ L each contained the appropriate primers: T7/T7-reverse for pET52-47, pGEX-forward/T7-reverse for pET49b as well as pGEX-6P1 and construct specific primers for pHMBP plasmids. Two to five single colonies were picked from the plate and dipped into individual tubes containing the reaction mixture before a PCR was run. PCR products were analysed by agarose gel electrophoresis.

2.2.11 Plasmid DNA purification, storage and sequencing

Single colonies with the correct size of inserts were used to inoculate LB (Luria-Bertani broth) media supplemented with the appropriate antibiotics. 5 mL media for miniprep or 200 mL for Maxiprep was used to grow over night at 37 °C shaking at 200 rpm. The QIAprep Spin Miniprep or the HiSpeed Maxiprep Kit (both Qiagen) respectively was used for plasmid purification following manufacturer's protocol and verified by sequencing (Source BioScience, LifeSciences or GATC Biotech). DNA was stored at -20 °C.

2.2.12 Site-directed mutagenesis

Primers have been designed containing the desired mutation to generate altered synthesized plasmids via PCR. Site-directed mutagenesis was performed using the QuikChange II site-directed mutagenesis kit (Agilent Technologies) including the PfuUltra High-Fidelity DNA polymerase (Table 2-7 and 2-8). This method is based on the fact that DNA isolated from almost all *E. coli* strains is methylated whereas the PCR-synthesized plasmids are not. Following thermal cycling, the parental DNA is digested using the endonuclease *Dpn* I, which specifically targets methylated DNA, leaving the mutation-containing synthesized DNA. 1 μ L of *Dpn* I-treated DNA was taken for transformation into XL1-Blue supercompetent cells (Agilent Technologies). All plasmids were verified by DNA sequencing as detailed above.

Reagent	Volume (μ L)
10x reaction buffer	5
dsDNA template (50ng/ μ L)	1
Forward primer (125ng/ μ L)	1
Reverse Primer (125ng/ μ L)	1
dNTP mix	1
ddH ₂ O	40
PfuUltra HF polymerase	1
Total Volume	50

Table 2-7 PCR reaction mixture for site-directed mutagenesis.

Step	Temperature ($^{\circ}$ C)	Time (sec)
1. Initial denaturation	95	30
2. Denaturation	95	30
3. Annealing	55	60
4. Extension	65	60/kb
Repeat steps 2-4 16 times		
5. Final extension	65	120
6. Cooling	10	300

Table 2-8 PCR cycling protocol for site-directed mutagenesis.

2.3 Protein expression and purification

2.3.1 Bacterial expression strains

For recombinant protein expressions, the plasmids were transformed into the widely used *E.coli* host strain BL21, with advantage of its Lon and ompT proteases deficiency that otherwise can cause protein degradation. Commercially available modified BL21 strains are often used depending on specific requirements for protein expression. Throughout this project, two specific strains were used: Rosetta2 (DE3) (Merk) and BL21 (DE3) gold (Agilent Technologies). The DE3 designation indicates that the bacterial strain contains a λ DE3 lysogen, which carries a chromosomal copy of the T7 RNA (ribonucleic acid) polymerase gene under control of the lacUV5 promoter and is inducible with lactose or the non-hydrolysable lactose analogue isopropyl β -D-1-thiogalactopyranoside (IPTG) (Generon). These strains are suitable for protein expression from target genes cloned into the T7 promoter containing pET vectors. Further, Rosetta2 (DE3) is modified with an additional chloramphenicol-resistant plasmid (pRARE2), which can enhance protein expression by supplying tRNAs for seven rarely used codons in *E.coli*: AGA, AGG, and CGA for arginine, GGA for glycine, AUA for isoleucine, CUA for leucine and CCC for proline. BL21 (DE3) gold strain on the other hand features the high transformation efficiency phenotype (Hte) and have the gene encoding endonuclease I (endA) inactivated, which otherwise degrades isolated mini-prep plasmid DNA.

2.3.2 Overexpression

2.3.2.1 Small-scale expression

Newly generated and purified plasmids were transformed into bacteria strains as mentioned above and usually first expressed on a small scale: A single colony was picked to inoculate a starter culture of 0.5 ml LB medium, containing the appropriate antibiotic resistance: 50 µg/mL kanamycin for pET49b and pHMBP, 50 µg/mL carbenicillin for pET52-47 and pGEX-6P1, 34 µg/mL chloramphenicol when using the Rosetta2 (DE3) strain. 10 µL starter culture was used to inoculate the main culture of 1 ml ZYM505 rich medium in sterile 48-deep well plates. The cultures were grown at 37 °C and 800 rpm to a mid log-phase OD₆₀₀ of about 6 and, if not stated otherwise, cooled down to 18 °C before inducing them with 1 mM IPTG, followed by an incubation time of about 16 h. Cells were harvested by centrifugation at 4000 rpm at 4 °C for 15 min, supernatant discarded and pellets stored at -80 °C. Novel constructs usually also included a screen for induction temperatures and induction time. 20 µL samples before and after induction were collected for Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis.

2.3.2.2 Large-scale expression

If not stated otherwise, the growth media used was LB supplemented with the appropriate antibiotics. Starter cultures of 10 - 25 mL growth media and inoculation of the main cultures was performed equally to small-scale expressions. Large-scale expressions varied between 0.75 and 12 L and were grown as 750 mL or 1 L cultures in sterile 2 L flasks in a shaking incubator at 37 °C, 210 rpm. The cultures were grown to a mid-log phase OD₆₀₀ of about 0.6-0.8 and cooled down to 18 °C before inducing them

with 1 mM IPTG, followed by an incubation over-night. Cells were harvested by centrifugation at 4000 rpm at 4 °C for 30 min, supernatant discarded and pellets stored at -20 °C.

2.3.3 Buffers and reducing agents

Buffers throughout this work were made with deionised water (MilliQ), filtered through a 0.22 µm membrane filter. Buffers for chromatography were also degassed. The pH of buffers was adjusted to keep it at ± 1 pH unit from the protein's pI (isoelectric point) to increase the charge on the protein and favour its interaction with the solvent. Depending on purification methods and experiments selected, the reducing agents DTT (dithiothreitol) (Thermo Scientific) or TCEP (tris(2-carboxyethyl) phosphine hydrochloride) (Thermo Scientific) were added to the buffers to prevent oxidation of cysteine residues. TCEP, compared to DTT, has the advantage to be resistant to air oxidation, is compatible with Ni-affinity chromatography and interferes less in UV spectroscopy applications.

2.3.4 Small-scale solubility screen

The protocol for small-scale protein expression and solubility screens has been developed by Vangelis Christodoulou (The Francis Crick Institute). The whole lysis process is carried out using the 48-well plate the pellets are stored in. The buffers used are listed in Table 2-9 for either His₆-tagged or GST-fusion proteins. Cell pellets were re-suspended in 0.5 mL re-suspension buffer by shaking the 48-deep well plate at 10 °C and 650 rpm for 15 min. 0.5 mL lysozyme containing lysis buffer 1 was added, followed by incubation on the shaker for another 15 min. A hypotonic shock was then induced

by adding 1 mL of H₂O (4 °C). After further 15 min, the plate was removed from the shaker and 1 mL of lysis buffer 2 was added. The deep-well plate was sealed with clear tape and the solution was mixed by just inverting 3 to 4 times. To complete the lysis process, the samples were incubated for 15 min at 4 °C before centrifugation at 5900 rpm for 1 h at 4 °C. 20 µL samples of the cleared lysate were taken for SDS-PAGE gel analysis, and the remaining lysate (approx. 3 mL) was loaded on a 96-deep well receiver plate containing 96 columns with 50 µL of either Ni-NTA beads (Qiagen) for HIS₆-tagged proteins or G-sepharose 4B beads (GE Healthcare) for GST-fusion proteins. The columns were washed with 4 x 0.5 mL of wash buffer, using a QIAvac vacuum manifold (Qiagen) before a further samples for analysis of the flow through was collected. Bound proteins were eluted by adding 60 µL of elution buffer and collected by a short centrifugation step. 15 µL of the eluted fraction was used for another analysis step *via* SDS-PAGE gel.

2.3.5 Cell lysis and sample preparation

Protein purification was usually carried out at 4°C. Cell pellets were thawed and re-suspended in 6 ml per gram cell pellet of lysis buffer (HIS₆-constructs; 50 mM Tris-HCl pH 8, 10 mM Imidazole, 300 mM NaCl and 0.5 mM TCEP. All other constructs; 50 mM Tris-HCl pH 8, 300 mM NaCl, 2 mM EDTA, 1 mM DTT).

	GST	HIS
Re-suspension Buffer	20 mM Tris-HCl pH 7.7	20 mM Tris-HCl pH 7.7
Lysis Buffer 1	20 mM Tris-HCl pH 7.7 40% succrose (w/v) 2 mg/mL lysozyme 2 mM EDTA	20 mM Tris-HCl pH 7.7 40% succrose (w/v) 2 mg/mL lysozyme 2 mM EDTA
Lysis Buffer 2	150 mM Tris-HCl pH 7.7 900 mM NaCl 0.45% OTG (w/v) 30% glycerol 30 mM MgSO ₄ 12 U/mL Dnase 1 3 mM TCEP 0.3 mM PMSF	60 mM Tris-HCl pH 7.7 120 mM Imidazole pH 8.0 900 mM NaCl 0.45% OTG (w/v) 30% glycerol 30 mM MgSO ₄ 12 U/mL Dnase 1 3 mM TCEP 0.3 mM PMSF
Wash Buffer	100 mM Tris-HCl pH 7.7 300 mM NaCl 1 mM TCEP	100 mM Tris-HCl pH 7.7 40 mM Imidazole pH 8.0 300 mM NaCl 1 mM TCEP
Elution Buffer	100 mM Tris-HCl pH 7.7 300 mM NaCl 10 mM reduced Glutathione 1 mM TCEP	100 mM Tris-HCl pH 7.7 500 mM Imidazole pH 8.0 300 mM NaCl 1 mM TCEP

Table 2-9 Lysis and purification buffers for small-scale expression and solubility screen.

Re-suspension was supplemented with 1 protease inhibitor cocktail tablet EDTA-free (Roche) per 50 mL lysate to prevent protein degradation. 10 U/ml of Benzonase® Nuclease (Novagen) and 0.25 mg/mL (end concentration) of hen's egg lysozyme (Sigma Aldrich) were added to reduce viscosity and support lysis. The samples were incubated at 4 °C for 30 min. To complete the lysis process, the lysate were sonicated on ice at 40 % power at 50 % duty cycle, 3-5 times for 30 sec with 1 min rest between cycles (Branson Sonifier 450). To clarify the lysate, samples were centrifuged at 20,000 rpm at 4 °C for 40 min in a Beckman JA-25.50. Samples of the lysate as well as the supernatant were analysed by SDS-PAGE to monitor the purification steps.

2.3.6 His-affinity chromatography

Ni²⁺-affinity chromatography was carried out as an initial purification step of His₆-tagged proteins. 1-2 mL Ni-NTA beads (Qiagen) per 40 mL clarified supernatant was added and samples incubated for 1 h at 4 °C before loading on a PolyPrep disposable column (BioRad). Beads were washed with 20 CV wash buffer (50 mM Tris-HCl pH 8, 20 mM Imidazole, 800 mM NaCl, 0.5 mM TCEP). The proteins were eluted by adding 5 CV elution buffer (50 mM Tris/HCl pH 8, 500 mM Imidazole, 300 mM NaCl). The eluted protein was concentrated for the next purification step using VivaSpin concentrators (Sartorius) with a molecular weight cut-off of 5 kDa. Samples for SDS-PAGE analysis were collected from the flow through, beads and elution. Ni²⁺-affinity chromatography was also used for on-column refolding of His-tagged proteins. Details are described in section 2.3.11.2.

2.3.7 GST-affinity chromatography

GST-fusion proteins were purified via GST-affinity chromatography by adding 1-2 mL G-sepharose 4B beads (GE Healthcare) per 40 mL clarified supernatant and incubated for 1 h at 4 °C before loading on PolyPrep disposable columns. Beads were washed with 20 CV lysis buffer buffer.

2.3.7.1 Elution without tag removal

The protein was eluted from the column by applying 10 mM reduced Glutathione (Merck). The eluted protein was concentrated for the next purification step and samples for SDS-PAGE analysis were collected from the flow through, beads and elution.

2.3.7.2 Elution including tag removal

After washing the beads (containing the bound protein) with lysis buffer, the beads were further washed with 5CV 3C-cleavage buffer (50 mM Tris-HCl pH 8, 800 mM NaCl, 1 mM DTT) for equilibration. The beads were then removed from the column and transferred into a 15 mL falcon tube with additional 3C buffer. Approximately 0.5 mg GST-tagged rhinovirus 3C protease (made in house) per 1 L culture was used to remove the GST-tag by incubating over night at 4 °C. Beads were transferred back to the column and the protein eluted with 3C buffer. The eluted protein was concentrated for the next purification step and samples for SDS-PAGE analysis were collected from the flow through, wash, beads – before 3C cleavage, beads – after 3C cleavage and elution.

2.3.7.3 GST-affinity purification of the modified tandem CARD construct of NOD2

A specially designed construct of the tandem NOD2-CARDs (amino acid (aa) 28-218, pGEX-6P1) containing a thrombin cleavage site between the two CARDs was used to express and purify the CARDs individually [131]. The protein was loaded on the column and washed as described above in 2.3.7. The beads containing the bound protein were equilibrated with a buffer suitable for thrombin cleavage (50 mM Tris-HCl, 150 mM NaCl, 2 mM CaCl₂, 2 mM DTT). 25U thrombin (Thermo Scientific) was added to cleave NOD2-CARDb over night at 4 °C. NOD2-CARDb was eluted using 3C-cleavage buffer because of the high salt content as well as to prepare for the following 3C-protease cleavage, which was carried out as described above.

2.3.8 Amylose-affinity chromatography

MBP-fusion proteins were purified via amylose-affinity chromatography by adding 1-2 mL amylose beads (NEB) per 40 mL clarified supernatant and incubated for 1 h at 4 °C before loading on PolyPrep disposable columns. Beads were washed with 20 CV lysis buffer. The protein was eluted from the column by applying 10 mM Mannose (Sigma Aldrich). The eluted protein was concentrated for the next purification step and samples for SDS-PAGE analysis were collected from the flow through, beads and elution.

2.3.9 Ion exchange chromatography

The tandem CARDS of NOD2 form a tight intramolecular interaction, which makes another purification step necessary for the NOD2-CARDa sample eluted after 3C cleavage to remove contaminating NOD2-CARDb. Remaining CARDb was removed on an anion exchange column at pH 7.5. This method is based on the fact that at this pH, CARDa with a pI of 4.80 has a higher net negative charge than CARDb, which has a pI of 6.71, and binds more strongly to the positively charged matrix. The NOD2-CARDa sample in 3C cleavage buffer was diluted to 75 mM NaCl and applied on an equilibrated HiTrap Q FF 5 ml column (GE Healthcare) using a flow rate of 5 mL/min and washed with 3 CV buffer. NOD2-CARDa was eluted by applying a step-wise salt gradient: 150 mM, 300 mM, 450 mM, 600 mM and 1M NaCl, using 1CV each with another final 3 CV of 1M NaCl. Fractions were analysed via SDS-PAGE.

2.3.10 Size exclusion chromatography

A final purification step was performed using size exclusion chromatography (SEC) to remove high-molecular contaminants, change buffer and to ensure mono-dispersity of the sample. The samples were concentrated to 4 mL and ejected onto either a pre-equilibrated Superdex-75 16/60 or Superdex-200 column (GE Healthcare), depending on the size of the protein. Columns were connected to an Äkta Prime (GE Healthcare) and typically run in 50 mM Tris-HCl pH 7.5 (or 50 mM HEPES pH 7.5), 150 mM NaCl, 2 mM DTT. Collected fractions were analysed on SDS-PAGE gels, pooled and concentrated.

2.3.11 Protein refolding

2.3.11.1 Batch semi-refolding with 3 M urea

The cell pellet was re-suspended in 50 mL lysis buffer (100 mM NaH_2PO_4 , 10 mM Tris-HCl, 3 M urea) at pH 8. One protease inhibitor cocktail tablet EDTA-free was added to the re-suspension. For lysis, the samples were sonicated at 40 % power and finally centrifuged at 20'000 rpm, 4 °C for 40 min. The supernatant was collected and 1 mL of Ni-NTA added before the sample was incubated for 30 min at 4°C. After loading the solution on a PolyPrep disposable column, a washing step with 12 CV wash buffer (3 M urea, 100 mM NaH_2PO_4 , 10 mM Tris-HCl) at pH 6.3 was carried out. The protein was eluted by decreasing the pH to 4.5. Before dialysis in 4 litres refolding buffer (1 M urea, 300 mM NaCl, 100 mM NaH_2PO_4 , 2 mM reduced glutathione), the eluted protein was diluted to <50 µg/mL. Dialysis was performed over night at 4 °C and the buffer was exchanged to 0 M urea on the next day. After 10 h, the pH was increased to 7 for the following concentration step.

2.3.11.2 On-column refolding with 6 M guanidinium hydrochloride (GdmCl)

Cell lysis was performed as described in 2.3.5. In this case, the supernatant was discarded and the pellet re-suspended with Pellet Wash Buffer 1 by extended vortexing. The centrifugation was repeated for 20 min, before the cell pellet was washed again with Pellet Wash Buffer 2, for a high salt washing step. During the washing steps, the pellet changed its colour to white. Finally, a third cycle of re-suspending and centrifugation was applied, using the Binding Buffer. At this step, the supernatant was collected and incubated with 1 mL Ni-NTA beads for 30 min at 4 °C. The solution is transferred to a disposable PolyPrep column and unbound protein was washed off using 10 CV Wash Buffer. Refolding was carried out by adding 10 CV wash buffer, containing just 1 M GdmCl and the second time 0 M. Protein was eluted with increasing imidazole to 500 mM.

Pellet Wash Buffer 1	Pellet Wash Buffer 2	Binding Buffer	Wash Buffer
50 mM Tris-HCl pH 8 100 mM NaCl 2 mM EDTA 1 % TritonX 100 1 mM TCEP	50 mM Tris-HCl pH 8 1 M NaCl 1 mM TCEP	50 mM Tris-HCl pH 8 300 mM NaCl 6 M GdmCl 5 mM Imidazole 1 mM TCEP	50 mM Tris-HCl pH 8 800 mM NaCl 6 M GdmCl 20 mM Imidazole 1 mM TCEP

Table 2-10 Buffers for on-column refolding.

2.3.12 NMR sample preparation

Protein expression and purification did not differ from the protocol described in this section 2.2 except that cultures were grown in either M9 minimal medium using ¹⁵N-ammonium sulfate (Cambridge Isotope Laboratories) as only nitrogen source or the commercially available Spectra9 medium (Cambridge Isotope Laboratories).

Purified protein samples were dialyzed over-night into NMR buffer (NOD2 constructs: 100 mM NaCl, 20 mM Na-Phosphate pH 7.1, 2 mM TCEP, all other proteins: 50 mM (or 100 mM) NaCl, 20 mM HEPES pH 7, 2 mM TCEP) and concentrated to 280 - 450 μ L before 5 - 10% (v/v) 99.9 %-deuterium oxide (Cambridge Isotope Laboratories) was added. 5 mm Norell® or Shigemi NMR tubes (both Sigma Aldrich) were used. Experiments have been kindly set up by Geoff Kelly, Biomolecular NMR Centre, The Francis Crick Institute. 1D ^1H - and 2D ^{15}N -HSQC spectra were recorded at 25 °C.

2.3.13 SDS-PAGE electrophoresis

Expression and purification steps were monitored by taking samples at each step for SDS-PAGE. In most cases, 20 μ L samples were mixed at a 1:1 ratio with 2x SDS loading buffer (Thermo Scientific), followed by heating up to 95°C for 5min. Pellet samples were first re-suspended in 8 M Urea. 5-20 μ L of the samples were loaded onto a pre-cast NuPage 4-12% BIS-TRIS Novex gel (Thermo Scientific) and run in 1x MES buffer (Thermo Scientific) at 200V. Typically Precision Plus Protein™ All Blue Standards Marker (BioRad) or SeeBlue®Plus2 (Thermo Scientific) was used as a molecular weight marker. Gels were stained with InstantBlue protein stain (Expedeon).

2.3.14 Determination of protein concentration

A NanoDrop ND-1000 spectrophotometer was used to determine protein concentration by ultraviolet (UV) absorption spectroscopy, in which the absorbance A between 220-350 nm was scanned. According to Beer-Lambert's law, the absorbance A is a linear function of the proteins concentration c (M):

Equation 2.1
$$A = \epsilon * c * l$$

The primary amino acid sequence of the protein, including any fusion tags, was used to calculate the theoretical 280 nm molar extinction coefficients ϵ ($M^{-1}cm^{-1}$) with the ProtParam calculator (see section 2.2.2). ϵ of a protein can be calculated making use of the average ϵ values for the three chromophores in proteins, which absorb light at 280nm: tryptophans, tyrosines and cysteines forming disulphide bonds, assuming that the extinction coefficients of the amino acids are the same, both in the protein and free in solution. Concentration c of the protein was then determined using molar absorption at 280 nm with a path length l of 1 mm. Amino acids with aromatic rings are the primary reason for the absorbance peak at between 276 and 282 nm, therefore the absorbance spectrum for a protein sample scanned between 250-350 nm (near-UV) mainly reflects tryptophan (Trp) and tyrosine (Tyr) contents as the main contributors and to a smaller extent phenylalanine (Phe) (Figure 2-2) [135].

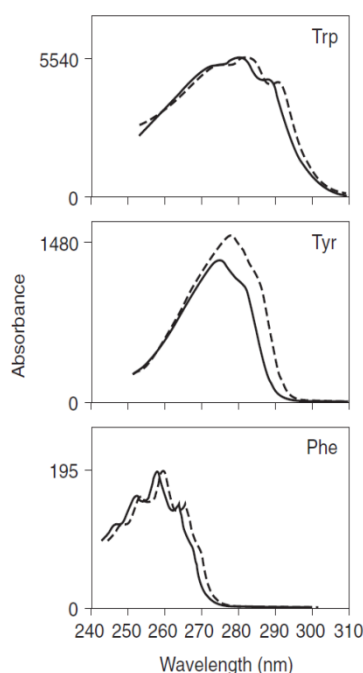


Figure 2-2 UV absorption spectra of Trp, Tyr and Phe. The three spectra illustrate aromatic model compounds (solid lines) versus “average” aromatic amino acid residues as determined using a set of globular, water-soluble proteins (dotted lines). Figure adapted from Current Protocols in Protein Science, unit 7.2 [135].

In addition to the protein concentration, a near-UV scan was also used to monitor aggregation and DNA contamination in protein samples. Aggregations cause light scattering that can be detected at wavelengths >300 as light is scattered away from the detector, resulting in a distorted spectrum. DNA contamination, on the other hand, is typically characterized by a “flattening out” of the spectrum at <280 nm as nucleic acids absorb light around 250-270 nm.

2.3.15 Concentration, verification and storage of protein samples

VivaSpin centrifugal concentrators with a molecular weight cut-off of 5 kDa were used for sample concentration according to manufacturer’s instructions. The purity of the protein sample was analysed by SDS-PAGE and molecular weight was verified by electrospray ionisation mass spectrometry (performed by Dr Steve Howell, The Francis Crick Institute). Final purified protein samples were aliquoted (20-50 µL), flash-frozen in liquid nitrogen and stored at -80°C.

2.4 Biochemical and biophysical analysis

2.4.1 Co-expression and GST-pull downs

Two CARD constructs of interest were co-transformed and -expressed in *E.coli*, followed by GST-pull down experiments to study direct CARD-CARD interactions. This experiments is based on the fact that the constructs are cloned into two expression plasmids encoding for different tags and antibiotic markers as further described below.

2.4.1.1 Small scale co-expression and purification screen of selected CARDS

CARD constructs have been cloned into the expression plasmids pET49b (N-terminal GST-tag and His₆-tag, kanamycin-resistance) and pET52-47 (N-terminal His₆-tag, ampicillin-resistance) to perform pull down experiments. Selected constructs have been co-transformed into BL21 (DE3) gold cells and co-expressed in 1 mL ZYM 505 media. Expression and following pull-down experiments have been performed as described in chapter 2.2.4.

2.4.1.2 NOD2-CARDab and RIP2-CARD

The expression plasmid pGEX-6P1 (N-terminal GST-tag, ampicillin-resistance) was used for NOD2-CARDab constructs and pHMBP (N-terminal His₆-tag and MBP-tag, kanamycin-resistance) for RIP2-CARD. The plasmids were simultaneously transformed into Rosetta2 (DE3) cells and grown in 150 mL LB media at 37 °C followed by induction at 18 °C over night. Cells were lysed in 25mL lysis buffer (300 mM NaCl, 50 mM Tris-HCl pH 8, 2 mM EDTA, 2mM DTT). After lysate clarification, 150 mL G-sepharose beads were added to the supernatant and incubated for 2h at 4°C. The sample was then loaded on a 5 mL gravitation column (Sigma Aldrich) and washed with 3x 1 mL lysis buffer. Proteins were eluted with 10 mM glutathione. Control experiment was performed by co-expression of the MBP-tag alone with NOD2-CARDab.

2.4.2 Quantitative analysis of protein-protein interactions

There are several ways to determine binding constants. Commonly used techniques are isothermal titration calorimetry (ITC), analytical ultracentrifugation, nuclear magnetic resonance (NMR), surface plasmon resonance or fluorescence spectroscopy.

The measured signal is related to the changes in concentration of free versus bound ligand, therefore, a protein-ligand interaction can be represented as:



The strength of the interaction can be described by either the association constant K_A or the dissociation constant K_D and are related as follow:

Equation 2.3
$$K_D = \frac{[P][L]}{[PL]} = \frac{1}{K_A}$$

According to the mass balance relationship, the total protein concentration (P_T) as well as the total ligand concentration (L_T) can be defined as:

Equation 2.4
$$P_T = [PL] + [P] \quad \text{and} \quad L_T = [PL] + [L]$$

Substituting Equation 2.4 into Equation 2.3 gives:

Equation 2.5
$$K_D = \frac{([P_T] - [PL]) ([L_T] - [PL])}{[PL]}$$

Solving Equation 2.5 for PL (concentration of formed complex):

$$K_D [PL] = P_T L_T - [PL] \cdot ([P_T] + [L_T]) + [PL]^2$$

thus

$$[PL]^2 - [PL] \cdot (P_T + L_T + K_D) + P_T L_T = 0$$

Solving the quadratic equation gives [PL]:

$$\text{Equation 2.6} \quad [PL] = \frac{(P_T + L_T + K_D) - \sqrt{((P_T + L_T + K_D)^2 - 4P_T L_T)}}{2}$$

This represents the general expression used to relate the measured signal obtained from the various quantitative techniques to the concentration of formed complex.

Differentiation of Equation 2.6 with respect to L_T gives:

$$\text{Equation 2.7} \quad \frac{d[PL]}{d[L_T]} = \frac{1}{2} + \frac{1 - (1 - r) / 2 - L_r / 2}{\sqrt{(L_r^2 - 2L_r - 2L_r (1 - r) + (1 - r)^2)}}$$

where $r = K_D/P_T$ and $L_r = L_T/P_T$

2.4.3 Nuclear magnetic resonance (NMR) spectroscopy

NMR spectroscopy was used to characterize the folded state of certain purified proteins. NMR spectra were obtained from several Bruker spectrometers with magnets ranging from 600MHz to 800 MHz and recorded at 25°C. Samples were dialysed into 50 mM HEPES pH 7.1, 100 mM NaCl, 2 mM TCEP and 10% D₂O was added to each sample. Final volumes were 500 µL for 5 mm Norell® or 300 µL for Shigemi

NMR tubes with protein concentrations varying depending on the constructs. NMR spectra and data analysis were kindly provided by Dr Geoff Kelly and Dr Kovilen Sawmynaden (The Francis Crick Institute).

2.4.4 Isothermal titration calorimetry (ITC)

ITC is a quantitative technique that can be used to study interactions (protein-protein, protein-ligand, protein-nucleic acid) and their thermodynamic properties (reviewed in [136-138]). ITC is a thermodynamic technique which monitors the heat change upon complex formation between molecules in solution. This is equal to the enthalpy change (ΔH) given that the pressure stays constant. Stoichiometry, association constant and enthalpy can be determined in a single experiment. The Gibb's free energy (ΔG°) and the entropy (ΔS°) of binding can then be determined from ΔH and K_A , according to the following relationships:

$$\text{Equation 2.8} \quad \Delta G^\circ = -RT \cdot \ln K_A$$

$$\text{Equation 2.9} \quad \Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$$

where ΔG° = Gibb's free energy (kcal mol^{-1}), R = universal gas constant ($1.987 \text{ cal K}^{-1} \text{ mol}^{-1}$), T = absolute temperature in Kelvin ($0^\circ\text{C} = 273.15 \text{ K}$), K_A = association constant (M^{-1}).

By performing ITC titrations at different temperatures, the change in heat capacity (ΔC_p°) for the reaction can be calculated from:

$$\text{Equation 2.10} \quad \Delta C_p^\circ = \frac{d\Delta H^\circ}{dT}$$

The change of binding heat capacity ΔC_p° is defined as the temperature dependence of the binding enthalpy change and reflects the change in solvation of the free molecules compared to the complex [139].

2.4.4.1 Determination of binding partners

The fraction of injected ligand ($v \cdot \Delta L_l$) which binds to the cell protein is proportional to the heat absorbed or developed during each calorimetric titration/injection (Q_l) as well as to the binding enthalpy of the reaction:

$$\text{Equation 2.11} \quad Q_l = v \cdot \Delta L_l \cdot \Delta H$$

where v = reaction volume of the cell.

Whereas probably most of the titrant will interact with the cell protein for the first few injections, the protein in the cell becomes more saturated as the titration carries on, therefore less heat is developed or absorbed and the size of the peaks decreases. After saturation, any heat observed represents the heat of dilution of the titrant. For the analysis, this heat of dilution can be corrected by averaging the enthalpies of the injections measured after reaching saturation and subtracting this average from the raw data. Alternatively, the heat of dilution can be obtained by performing a separate titration experiment in which the heat of dilution is measured by injecting the titrant into buffer only. The obtained binding data are then integrated and analysed using the

ORIGIN software, which is provided by the manufacturer (MicroCal). The heats of binding are normalised as a function of ligand concentration (kcal/mole of injectant). Further, the binding isotherm can be obtained by plotting the normalised heat per injection versus the molar ratio of ligand over protein in the sample cell ($[L_T]/[P_T]$). This can be fitted using a non-linear least-square fitting. The sigmoidal curve obtained allows then the determination of the binding parameters ΔH , K_A and n as shown in Figure 2-3.

2.4.4.2 Considerations and parameters in ITC

A number of parameters should be considered regarding sample preparation and experimental setup. Firstly, since ITC measures the total heat change in the sample cell during titration, both cell sample and titrant need to be dialyzed into the same buffer in order to minimize any artefacts from mismatched buffer components. A pH with low heat of ionization should be chosen so that the measured enthalpy does not reflect both complex formation and buffer ionization. Secondly, ΔH is temperature dependent and if no binding heat signal is detected, this could be due to a binding heat change near zero at a given temperature. In this case, experimental temperature should be changed by at least 5 to 10 °C in either direction. Additionally, an accurate determination of the protein concentration is crucial to determine the binding parameters, since the binding isotherm is obtained by plotting the normalised heat per injection versus the molar ratio of ligand over protein in the sample cell. To determine an optimal cell sample concentration ($[P]$), the unitless parameter c is used which characterises the shape of the binding isotherm (Equation 2.12) [140].

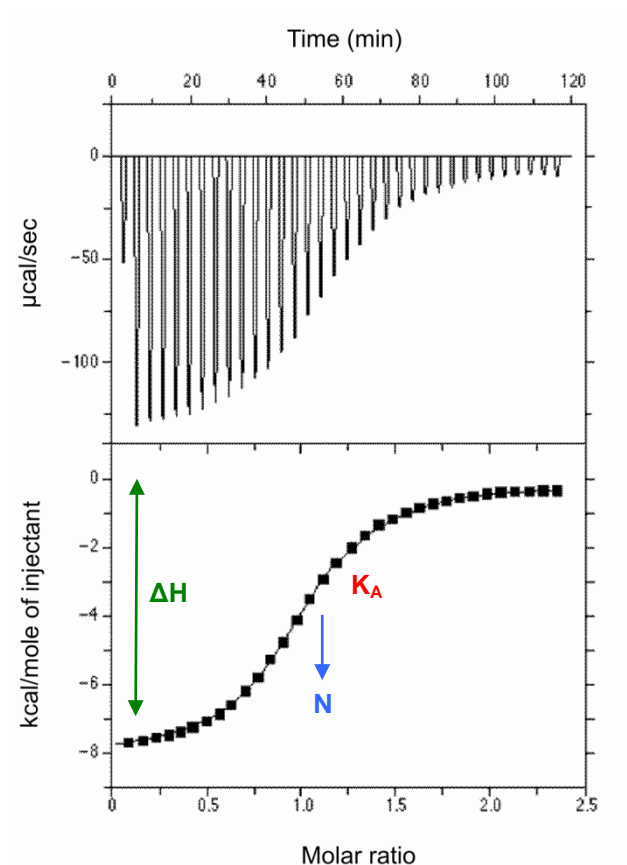


Figure 2-3 Example of an ITC titration. The upper plot shows the signal directly recorded of the power ($\mu\text{cal/sec}$) versus time. The lower plot shows the integrated heat plot; the amount of heat measured at each injection which is normalized to the number of moles of injectant (kcal/mol) versus the molar ratio of the injectant added to the cell sample. Figure adapted from Current Protocols in Protein Science, unit 20.4 .

The recommended c value lies between 10 and 100 in order to obtain a sigmoidal curve that enables accurate determination of K_A , ΔH and n (Figure 2-4).

Equation 2.12
$$c = K_A \cdot [P] \cdot n$$

where K_A = association constant, P = cell sample concentration and n = stoichiometry.

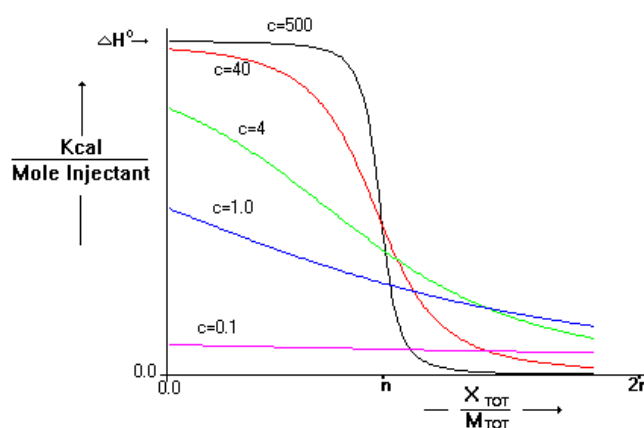


Figure 2-4 Simulated binding isotherms at various c values. The binding isotherm is characterized by the unitless value c , which reflects the shape of the curve: at low c values (<1), the binding isotherm is too shallow to allow reliable determination of K_A , n and ΔH while at high c values (>500), the binding isotherm is too steep and too few data points are collected at the transition point to determine K_A . A c value between 10 and 100 is recommended. Figure adapted from the MicroCal ITC200 user-manual, 2009.

2.4.4.3 Materials and methods

An ITC200 MicroCalorimeter (MicroCal) was used for all ITC experiments. Samples were dialysed into ITC buffer (50 mM HEPES-HCl pH 7.5, 150 mM NaCl, 2 mM TCEP). Typically, the protein concentration in the sample cell was 30-50 μM and 8 to 10-fold higher in the syringe. Titrations were carried out with 19 consecutive 2 μL injections at a stirring speed of 1000 rpm and with a duration of 4 seconds, spacing of 180 seconds and filter period 5 seconds. The first injection was always smaller (0.2 μL volume in 0.4 seconds) to compensate for dilutions and/or losses from the tip of the syringe during experimental set-up. Data from the first injection was not included in the analysis. Heats of dilution were determined from the last few injections of the titration if full saturation was reached and subtracted from the raw data. ITC data were analysed

using the ORIGIN7 software (MicroCal) supplied by the manufacturer and fitted by least-square procedures assuming a one-site binding model.

2.4.5 Far-UV circular dichroism

CD is an absorption spectroscopy technique used to gain secondary structure information of proteins in solution (reviewed by Martin and Schilstra [141]). This technique is based on the absorption of circularly polarized light by molecules in an asymmetric environment, where the difference in absorption of left and right circularly polarized light is measured. These follow the Beer-Lambert law (Equation 2.1) and the observed parameter ΔA can be defined as:

$$\text{Equation 2.15} \quad \Delta A = \Delta A_L - \Delta A_R = (\epsilon_L - \epsilon_R)C \cdot l = \Delta \epsilon \cdot C \cdot l$$

where $\Delta \epsilon$ = differential molar extinction coefficient, c = sample concentration, l = cuvette path-length, A_L and A_R = absorption of left and right circularly polarized light.

Apart from the protein secondary structure content, this technique can also be used to gain information about the stability of proteins by monitoring CD signals on a fixed wavelength (e.g. 222 nm) at variable temperatures.

2.4.5.1 General considerations

A critically important thing is that the instrument must always be purged with high-purity, oxygen-free, nitrogen (generally run at ~ 3 -5 L/min) for at least 20 minutes before starting the light source and throughout the measurements. If oxygen is

present it may be converted to ozone by the far-UV light from the high intensity arc, and ozone will damage the expensive optical surfaces.

As with all spectroscopic techniques the samples should be of the highest possible purity. Misleading results can be obtained even with relatively low levels of impurities if these have strong CD signals. Sample concentrations should be accurately known; this is particularly essential for the analysis of far-UV CD spectra for secondary structure content.

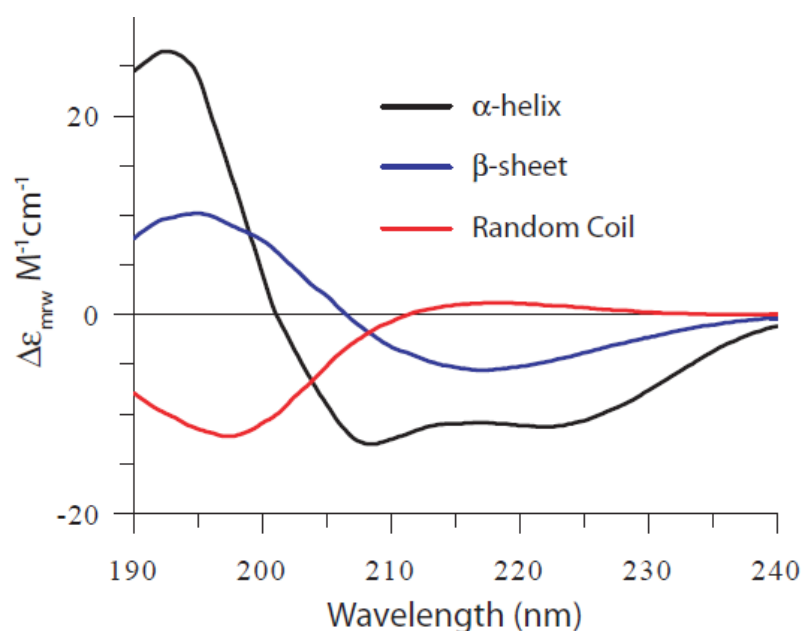


Figure 2-5 CD reference spectra. Far-UV CD spectra illustrating the characteristic curves for α -helix (black), β -sheet (blue) and random coil (red) conformations. These were obtained from the analysis of poly-lysine under various conditions. Figure adapted from unpublished data with kind permission from Dr Steve Martin, The Francis Crick Institute.

One major problem in CD measurements is that the signals become seriously distorted if too little light reaches the photomultiplier. In far-UV measurements, the absorbance of the sample itself is generally rather small and the major problems are caused by buffer components. The majority of simple buffer components will generally permit CD measurements to measure below 200 nm. However, high concentrations (> 1mM) of reducing agents (dithiothreitol and 2-mercaptoethanol) should be avoided whenever possible.

2.4.5.2 Materials and methods

CD analysis has been carried out with a protein concentration of 0.15 mg/mL in a 1 mm path length quartz cuvette in CD buffer (150 mM NaCl and 25 mM Tris-HCl pH 8). CD spectra were obtained on a JASCO-J815 spectropolarimeter. All spectra were corrected for buffer signals. Far-UV CD was used to study secondary structure content and spectra were collected from 260 nm to 195 nm at 20 °C, with typically 15 averaged scans. Thermal unfolding was carried out at 222 nm at increasing temperatures from 10 °C to 95 °C. CD data analysis was performed with help from Dr Steve Martin and Dr Laura Masino (The Francis Crick Institute). The CD signal was measured in millidegrees and converted to $\Delta\epsilon_{mrw}$ using equation 2.17:

$$\text{Equation 2.17} \quad \Delta\epsilon_{mrw} = \frac{S \cdot mrw}{32980 \cdot c \cdot l}$$

where $\Delta\epsilon_{mrw}$ = mean residue CD extinction coefficient ($M^{-1}cm^{-1}$), S = CD signal in millidegrees, mrw = mean residue weight of the protein (molecular weight of protein divided by its number of residues), c = protein concentration (mg/mL) and l = cell path length (cm).

2.4.6 Thermofluor

This fluorescence based experiment was used in this work solely to search for optimal buffer conditions or additives under which the protein sample showed maximum thermal stability. The thermofluor assay is based on the concept that a hydrophobic fluorophore is added to the sample prior temperature increase: The probe is quenched in aqueous solution, however, when the protein undergoes thermal unfolding, the hydrophobic regions of the protein become exposed and the probe will be able to bind those regions. The resulting increase in fluorescence emission can be plotted as a function of temperature [142].

2.4.6.1 Materials and methods

5 μ L protein solution at 1 mg/mL was mixed with the prior diluted (1:300) fluorophore SyPro Orange (Thermo Scientific) and added to 96-well plates containing 12.5 μ L of various buffers obtained from commercially available screens: Solubility & Stability and Slice pH both from Hampton Research. Plates were sealed and heated from 20 to 90 °C in increments of 1 °C/min in a Real time PCR machine (QPCR, Applied Biosystems).

The data were analysed using Prism software, where the sigmoidal curve was fitted using the Boltzmann equation in order to determine T_m .

2.4.7 SEC MALLS

Size-exclusion chromatography coupled to multi-angle laser light scattering (SEC-MALLS) provides a technique which allows the determination of the molecular weight of your protein sample. MALLS represents a type of static light scattering and is based on the principle that the intensity of scattered light is proportional to the molar

mass of a particle and that this intensity also depends on the scattering angle. For this reason MALLS data is collected at several angles to the incident laser beam [143].

2.4.7.1 Materials and methods

For SEC-MALLS analysis, samples were concentrated to values between 1.5 to 3 mg/mL and applied to a Superdex-200 10/300 GL column (GE Healthcare) mounted on a Jasco HPLC pre-equilibrated in MALLS buffer (CARD9: 100 mM BIS-TRIS propane pH 7.5, 250 mM LiCl, 2 mM TCEP, 3 mM Sodium azide. Others: 150 mM NaCl, 50 mM HEPES pH 8.0, 2 mM TCEP, 3 mM Sodium azide). The system was run at a flow rate of 1 mL/min. Scattered light intensity of the column eluate was recorded at 16 angles using a DAWN-HELEOS laser photometer (Wyatt Technology). Average molecular masses and poly-dispersity terms were determined using the ASTRA software version 5.1 (Wyatt Technology). Data analysis was performed with help from Dr Ian Taylor and Dr Neil Ball (The Francis Crick Institute).

2.4.8 Fluorescence polarisation assays (FPA)

Polarisation measurements can provide information on the size and shape of proteins. These measurements are based on the principle of photoselective excitation of fluorophores by polarized light. In most cases, the fluorophores are oriented randomly in the solution. Excitation with polarized light will result in a selective excitation of those fluorophore molecules whose absorption transition dipole is parallel to the electric vector of the excitation. This selective excitation results in a partially oriented population which can have polarized fluorescence emission. The extent to which the emitted light will be polarized depends on how fast the fluorophore tumbles in

solution. If the fluorophore tumbles fast (small molecule) during the fluorescence lifetime, then the emitted light will be depolarized. If it tumbles slowly, as when attached to a large protein, then the emitted light will be polarized. Fluorescence polarization (P) is the ratio of orthogonal fluorescence intensities defined by equation 2.18:

$$\text{Equation 2.18} \quad P = \frac{I_{VV} - G \cdot I_{VH}}{I_{VV} + G \cdot I_{VH}}$$

where I_{VV} and I_{VH} are the vertical and horizontal emission components, respectively, with the exciting beam polarized in the vertical direction. The instrument factor G corrects for polarization introduced by the optical components of the detection system [144].

2.4.8.1 Instrumentation

A simple representation of a general spectrofluorimeter is shown in Figure 2-6. It consists of a light source which typically is capable of exciting light ranging from 200 to 900 nm (e.g. xenon lamp). This is followed by an excitation monochromator, an optical device which transmits a narrow band of light wavelengths onto the sample. Fluorescence emitted from the sample cell enters the emission monochromator, which is positioned at a 90° angle to the excitation light path to eliminate background signal and reduce noise due to stray light. Both monochromators normally contain adjustable slits at the entrance and exit to further control the range of transmitted light. The emitted light enters the photomultiplier tube, which amplifies the signal and provides a voltage output proportional to the emitted light intensity [144].

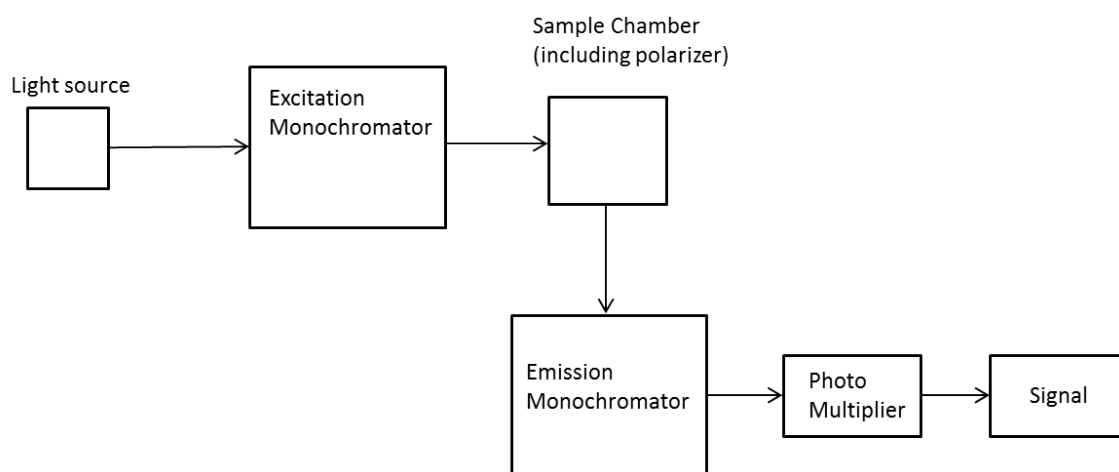


Figure 2.6 Schematic diagram of a general spectrofluorimeter.

2.4.8.1 Materials and methods

Labelling

Alexa-488-C5-maleimide (Thermo Scientific) was dissolved in dimethyl sulfoxide to obtain a stock solution of around 3 mM. Labelling of the proteins was carried out in 500 mM NaCl (or 150 mM), 20 mM Tris-HCl at pH 7.5 and 0.5 mM TCEP. Purified His₆-MBP-BCL10 C29A/C57A and His₆-MBP-RIP2-CARD C455A were mixed with 3-fold molar excess of Alexa-488-C5-maleimide and incubated at room temperature for 2 hours. Gel filtration chromatography (Superdex-200 16/60) was used to remove the excess dyes and change buffer to 150 mM, 20 mM Tris-HCl at pH 7.5 and 0.5 mM TCEP.

Fluorescence polarization assay

Fluorescence polarization was measured using excitation/emission wavelengths of 490 nm/522 nm on a Jasco FP-8500. 80 µL of sample was loaded into a cuvette with 3 mm pathlengths. Instrument settings were as follows: slit excitation/emission; 5/5 nm,

sensitivity; low, response; 1 second, data interval; 10 seconds, duration; 7200 seconds. Fluorescence polarization assays were performed at 25 °C with a protein concentration of 10 μ M. 3 μ M CARMA1, CARD9 or NOD2-CARDab was added in selected experiments. 5 μ g 3C protease was added to the samples for removal of the MBP-tag after about 100 seconds.

Chapter 3

In search of soluble, interacting CARD pairs

3.1 Overview

To fully understand the mechanisms underlying recruitment of downstream effector molecules and signal transduction in immune signalling pathways, it is essential to investigate the protein-protein interactions involved on the molecular level. CARDs belong to the death domain superfamily and mediate many protein-protein interactions in the innate as well as the adaptive immune system. However, at the start of this project, only the structure of a 1:1 complex between Procaspase9-CARD and APAF1-CARD provided an insight into homotypic CARD-CARD interactions [116]. On the other hand, death domains have been shown to form multi-protein signalling platforms and are able to accommodate six binding partners simultaneously. Based on the apoptosome model, it has been suggested that CARDs are forming a ring-like structure. However, in this case, an additional oligomerisation domain (as provided by the NOD) is necessary to bring the CARDs into proximity. The question remains now, what are the general requirements for CARDs to form a complex and are these only with a 1:1 stoichiometry or are oligomeric signalling platforms also formed by CARDs?

A main reason for the lack of structural information so far is due to the tendency of CARDs to aggregate when attempted to purify them as individual domains. The aim of this chapter was therefore to screen all human CARDs in co-expression and co-purification (pull-down) experiments with the goal to find soluble CARD-pairs, which can be taken further for biochemical and biophysical investigations. The reason for using co-expression and pull-down experiments is firstly that the solubility of CARDs may be enhanced by forming complexes and therefore burying hydrophobic patches. Secondly, with the structure of the apoptosome in mind, beads would potentially mimic an additional oligomerisation domain by bringing CARDs into

proximity to form the platform possibly needed to pull down a binding partner. This way, we would screen for soluble, interacting CARD pairs addressing all possibilities mentioned above how CARDS may interact.

Excluded from this screen are the CARDS of cIAP1 and cIAP2; we are interested in interacting CARD pairs, but protein interaction studies by Dr Pascal Meier (ICR, UK) using the yeast two-hybrid system had not revealed any other CARD interacting with cIAP1 or cIAP2.

3.2 Expression and solubility screens of all human CARDS

As a first step, all human CARDS were cloned into two different expression vectors, which were meeting the requirements for co-expressions: The CARDS were cloned to be expressed as His₆-tagged as well as GST-fusion proteins. To control first, if these constructs are suitable for the co-expression experiments, the CARDS were tested for their expression and solubility individually.

3.2.1 Construct design and cloning

CARD constructs were designed using bioinformatics tools for secondary structure and domain boundary predictions (e.g. Jpred [145]), as the primary sequence similarity of the members of the DD superfamily is very low (4.7 - 25.3 %) [117]. A multiple sequence alignment of all 33 human CARDS generated using ClustalOmega is shown in Figure 3-1 [146]. The constructs were designed assuming that all CARDS contain a six helical bundle as shown in the structures of APAF1 and Procaspase-9 [116].

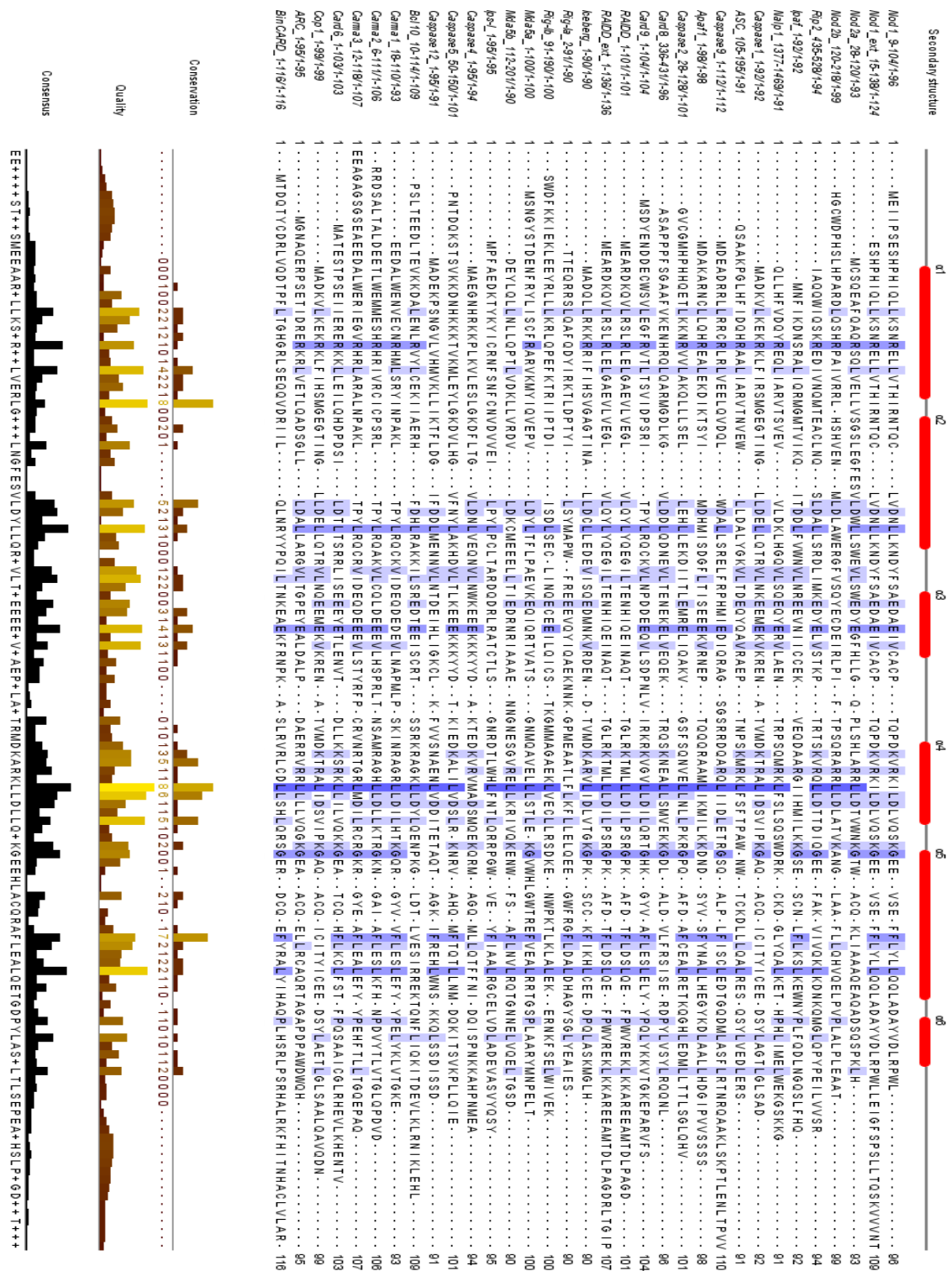


Figure 3-1 Multiple sequence alignment of all human CARDs. This alignment depicts sequence conservation of all 33 human CARDs, which is highlighted in blue. This Figure was generated using only the CARD sequences of all CARD-containing proteins. JalView was used including the alignment tool Muscle (<http://www.ebi.ac.uk/Tools/msa/muscle/>). Secondary structure elements were added based on the crystal structures of APAF1-CARD (PDB: 1CY5 and 4LWD respectively).

However, as domain boundaries were not obvious in all cases, the N- and C-termini were set generously and prolines were chosen in most cases as the indicators for flexible regions and unlikely to be involved in forming an alpha helix.

Based on published studies on NOD1 and RAIDD, two constructs have been tested each, one construct containing just the CARD, and another which was C-terminally extended; extended constructs of NOD1 and RAIDD have been shown to increase solubility and stability of the molecules. Additionally, for tandem CARDS, constructs containing both domains as well as the single domains have been designed.

Generating such a large number of recombinant proteins requires an efficient method for cloning. One suitable method is ligation-independent cloning (LIC), which does not require the use of restriction endonucleases, DNA ligase or alkaline phosphatases. Therefore, all constructs were made by LIC based PCR cloning into *E.coli* expression vectors using human cDNAs of CARD containing proteins as templates for the PCR (received from Pascal Meier [ICR, UK], Margot Thome [UNIL, CH] and SourceBioScience).

The vectors used were pET49b (N-terminal His₆-GST-tag, kanamycin resistance), and pET52-47 (N-terminal His₆-tag, ampicillin resistance) with the intention for the co-expression experiments to use the His₆-GST-fusion constructs to pull down the His₆-CARDS via GST-binding affinity chromatography using G-sepharose beads (GE Healthcare). The full description of the constructs can be found in the section in Table 3-1 and 3-2.

3.2.2 Small scale expression and solubility screen

In total 36 His₆-CARD (pET52-47) and 36 His₆-GST-CARD (pET49b) constructs have been successfully cloned. The constructs were transformed into BL21 (DE3) Gold cells and expressed in a small scale, using 48- and 96-deep well plates. To select the best temperature for protein expression, initial screens were performed with the majority of the constructs to compare the expression, yield and solubility of the proteins when incubated at 30 °C for 4 hours or 25 °C and 18 °C overnight. Similarly, IPTG concentrations of 0.1 mM and 1 mM were compared. In both cases, no significant differences were observed (data not shown), therefore, an induction temperature of 18 °C and an IPTG concentration of 1 mM was used for all following experiments.

The purification protocol used was provided by Vangelis Christodoulou (The Francis Crick Institute) and is shown schematically in Figure 3-2. The cells were lysed with lysozyme and the supernatant clarified before being applied to columns containing either G-sepharose or Ni-NTA beads for metal affinity chromatography. The flow through was collected for further analysis before the bound proteins were eluted. To monitor the purification, samples for SDS-PAGE analysis were collected from different purification steps: uninduced and induced cells, supernatant, flow through and elution. Samples of the beads could not be analysed in this set-up due to the use of 96-deep well plates. The resulting gels of the induced cells sample together with the elution are shown in Figure 3-3 for all the His₆-proteins and in Figure 3-4 for the His₆-GST-fusion proteins with Tables 3-1 and 3-2 containing the construct details. Gels containing the uninduced cell samples, supernatant and flow through are not shown.

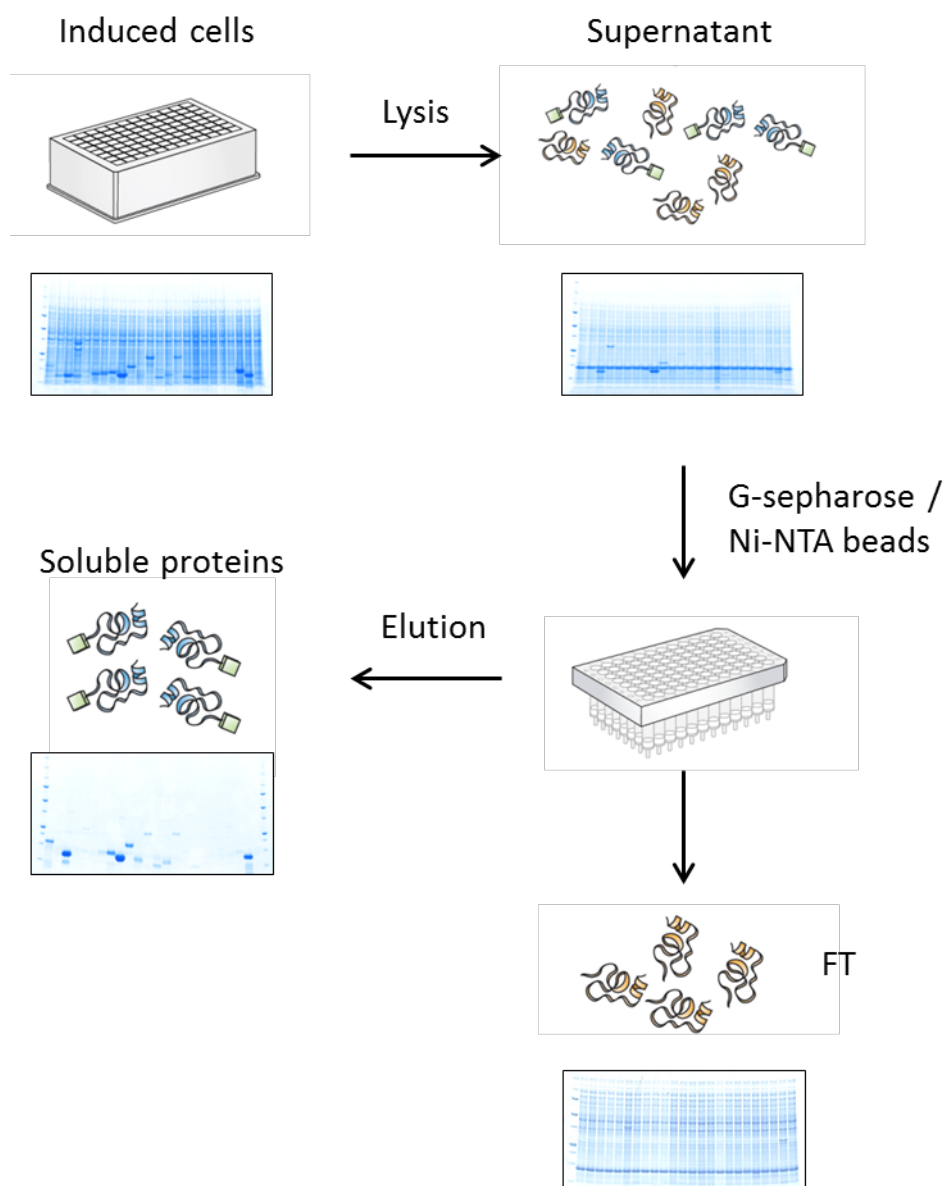


Figure 3-2 Small scale expression and purification. Small scale expression and purification experiments have been performed using 48- and 96-deep well plates, respectively. Induced cells were grown over night before being harvested. Cells were lysed and supernatants collected and clarified prior to loading on the columns containing either G-sepharose or Ni-NTA beads. The flow through (FT) was collected only for analysis before the bound proteins were eluted. SDS-PAGE analysis has been performed by taking samples of the un-induced and induced cells, the supernatants, the flow through and the eluted proteins.

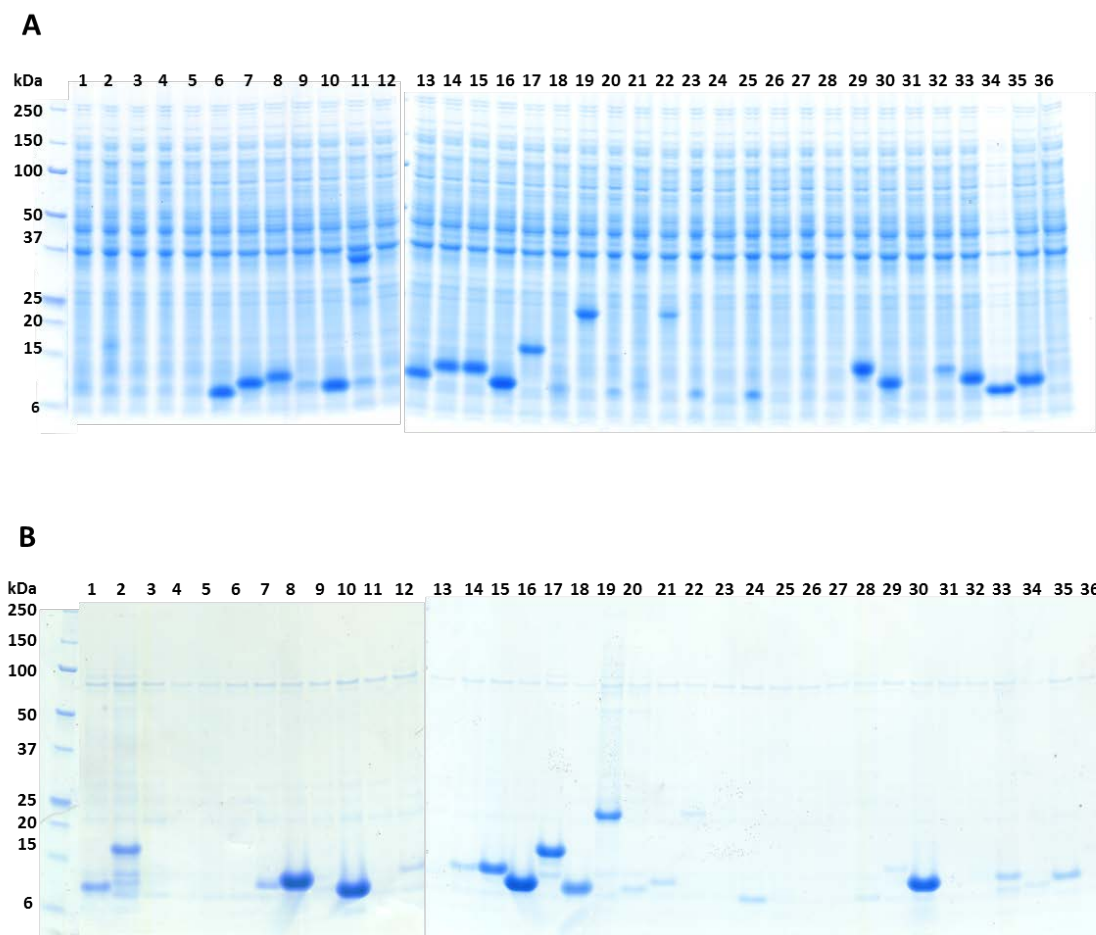


Figure 3-3 Small scale expression and solubility screen of His₆-tagged CARD constructs. SDS-PAGE analysis of the samples collected from the induced cells (A) and from the eluted proteins from the Ni-NTA columns (B). The constructs are numbered as listed in Table 3-1.

#	CARD	Construct	M _w (kDa)	pI	expressed	soluble
1	NOD1	9-104	13.3	5.62	✓	✓
2	NOD1 ext	15-138	16.4	6.09	✓	✓
3	NOD2 ab	28-218	23.8	5.97	✗	✗
4	NOD2 a	28-120	12.7	5.6	✗	✗
5	NOD2 b	120-218	13.5	6.66	✗	✗
6	IPAF	1-92	13	6.69	✓✓✓	✗
7	ASC	105-195	12.7	6.7	✓✓✓	✓
8	APAF1	1-98	13.4	6.17	✓✓	✓✓✓
9	RIP2	435-528	13.1	6.28	✓	✗
10	NALP1	1377-1469	12.9	7.99	✓✓✓	✓✓✓
11	Caspase1	1-92	12.5	6.28	✓✓	✗
12	Procaspase9	1-112	15.2	7.06	✓	✓
13	Procaspase2	28-128	13.6	6.44	✓✓✓	✗
14	CARD8	336-431	13.1	5.75	✓✓✓	✓
15	CARD9	1-104	14.2	6.14	✓✓✓	✓✓
16	RAIDD	1-101	13.7	5.76	✓✓✓	✓✓✓
17	RAIDD ext	1-136	17.7	5.88	✓✓	✓✓
18	Iceberg	1-90	12.3	6.48	✓	✓✓
19	RIG-I ab	2-190	24.6	5.47	✓✓✓	✓✓
20	RIG-I a	2-91	12.8	5.38	✓	✓
21	RIG-I b	91-190	14.1	6.3	✓	✓
22	MDA5 ab	1-201	25.3	5.28	✓	✗
23	MDA5 a	1-100	13.7	6.64	✓	✗
24	MDA5 b	112-201	12.6	5.1	✗	✓
25	MAVS	1-95	13.3	6.12	✓	✗
26	Caspase4	1-95	13.1	7.11	✗	✗
27	Caspase5	50-150	14	9.35	✗	✗
28	Caspase12	1-95	12.4	5.61	✗	✗
29	BCL10	10-114	15.1	7.14	✓✓✓	✗
30	CARMA1	18-110	13.2	5.86	✓✓✓	✓✓✓
31	CARMA2	6-111	14.3	6.58	✗	✗
32	CARMA3	12-118	14.6	5.73	✓	✗
33	CARD6	1-103	14.1	6.35	✓✓✓	✓
34	COP1	1-99	13.2	5.75	✓✓	✗
35	ARC	1-95	12.7	5.72	✓✓✓	✓
36	BINCARD	1-116	15.9	8.28	✗	✗

Table 3-1 His₆-CARD constructs from the small scale expression and purification experiment.

Listed are all His₆-CARD constructs and their details together with the results following the small scale expression and purification experiment as seen in Figure 3-3. Results are illustrated as heavily ✓✓✓, moderately ✓✓ and minimally ✓ expressed/soluble constructs based on the visual analysis of the SDS-PAGE gels. Unexpressed/insoluble constructs are marked with ✗.

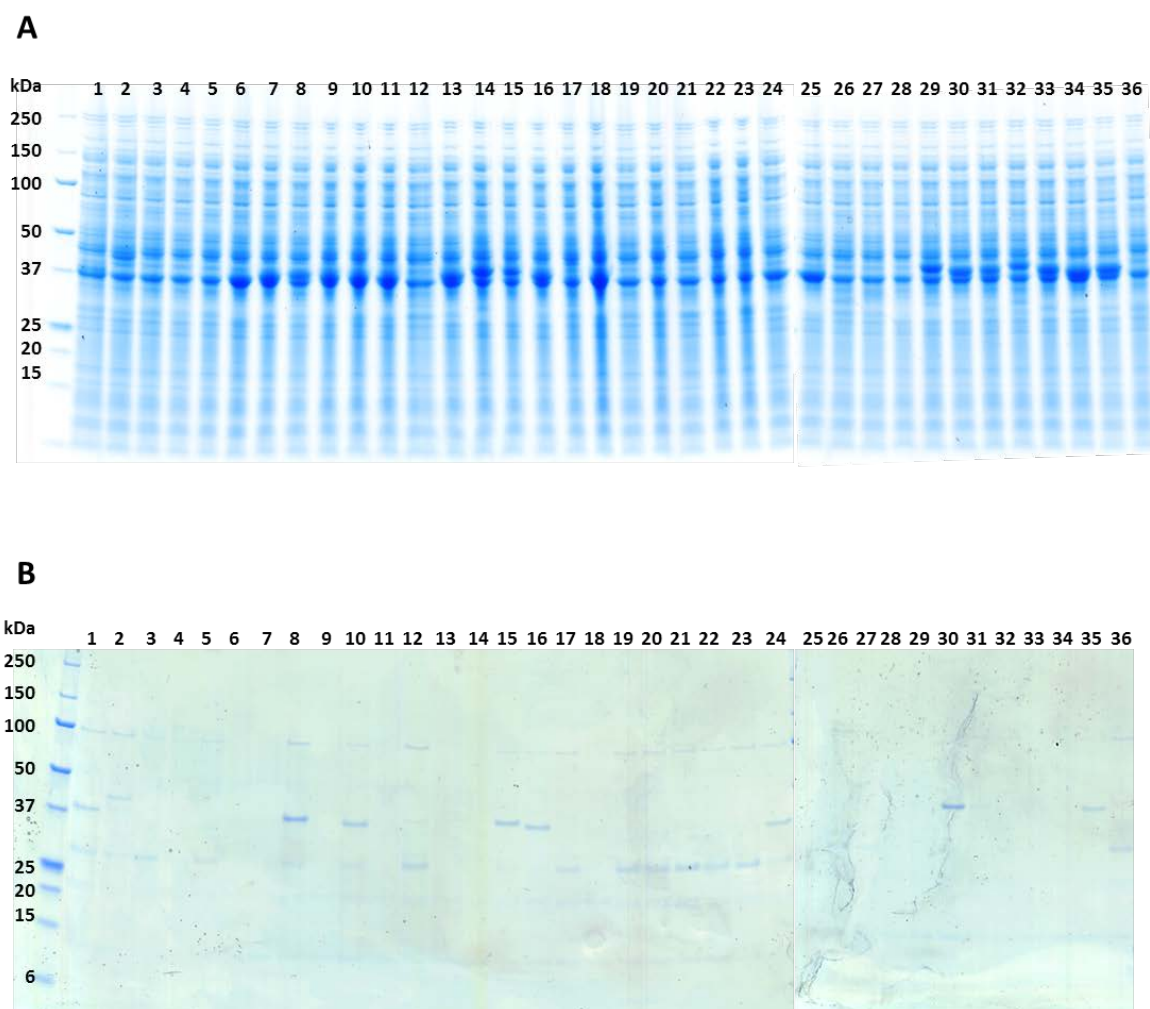


Figure 3-4 Small scale expression and solubility screen of GST-tagged CARD constructs using pET49b. SDS-PAGE analysis of the samples collected from the induced cells (A) and from the eluted proteins from the G-sepharose columns (B). The constructs are numbered as listed in Table 3-2.

#	CARD	Construct	M _w (kDa)	pI	expressed	soluble
1	NOD1	9-104	38.6	5.54	✓	✓
2	NOD1 ext	15-138	41.8	5.87	✓	✓
3	NOD2 ab	28-218	49.2	5.83	✗	✗
4	NOD2 a	28-120	38.1	5.53	✗	✗
5	NOD2 b	120-218	38.8	6.28	✗	✗
6	IPAF	1-92	38.3	6.17	✓✓✓	✗
7	ASC	105-195	38	6.17	✓✓✓	✗
8	APAF1	1-98	38.7	5.91	✓✓	✓
9	RIP2	435-528	38.4	5.89	✓✓✓	✗
10	NALP1	1377-1469	38.3	6.61	✓✓✓	✓
11	Caspase1	1-92	37.8	5.9	✓✓✓	✗
12	Procaspase9	1-112	40.5	6.28	✓	✗
13	Procaspase2	28-128	39	6.13	✓✓✓	✗
14	CARD8	336-431	38.4	5.61	✓✓✓	✗
15	CARD9	1-104	39.6	5.84	✓✓✓	✓
16	RAIDD	1-101	39.1	5.61	✓✓✓	✓
17	RAIDD ext	1-136	43	5.7	✓✓	✗
18	Iceberg	1-90	37.7	6.09	✓✓✓	✗
19	RIG-I ab	2-190	49.9	5.45	✓	✗
20	RIG-I a	2-91	38.1	5.41	✓✓	✗
21	RIG-I b	91-190	39.4	5.92	✓	✗
22	MDA5 ab	1-201	50.6	5.34	✓✓	✗
23	MDA5 a	1-100	39.1	6.11	✓✓	✗
24	MDA5 b	112-201	37.9	5.26	✓	✓
25	MAVS	1-95	38.6	5.82	✓✓	✗
26	Caspase4	1-95	38.5	6.34	✗	✗
27	Caspase5	50-150	39.4	8.1	✗	✗
28	Caspase12	1-95	37.7	5.53	✗	✗
29	BCL10	10-114	40.4	6.4	✓✓	✗
30	CARMA1	18-110	38.5	5.68	✓✓✓	✓
31	CARMA2	6-111	39.7	6.2	✓✓	✗
32	CARMA3	12-118	39.9	5.61	✓✓	✗
33	CARD6	1-103	39.4	6.04	✓✓	✗
34	COP1	1-99	38.6	5.61	✓✓✓	✗
35	ARC	1-95	38.1	5.72	✓✓	✓
36	BINCARD	1-116	41.3	8.28	✓	✗

Table 3-2 GST-CARD constructs from the small scale expression and purification experiment.

Listed are all GST-CARD constructs and their details together with the results following the small scale expression and purification experiment as seen in Figure 3-4. Results are illustrated as heavily ✓✓✓, moderately ✓✓ and minimally ✓ expressed/soluble constructs based on the visual analysis of the SDS-PAGE gels. Unexpressed/insoluble constructs are marked with ✗.

The screens clearly show that His₆-constructs are yielding more overexpressed protein overall as well as more soluble constructs in this small scale experiment. This is surprising as the GST-tag has been reported to improve protein solubility [147]. The GST-constructs seem to be expressed in most cases, although the SDS-PAGE analysis of the elution samples after GST-binding affinity purification results in rather weak bands. A further reason for using a GST-tag was that the purification using G-sepharose reduces non-specific binding when compared with Ni-NTA and hence would result in purer proteins when used for pull-down experiments. However, considering that the GST-tag did not show the expected increase in solubility, it was assumed that pulling down potential complexes on Ni-NTA bead would probably be more successful. As the pET-49b expressed GST-constructs contain an additional N-terminal His₆-tag, all GST-CARD constructs had to be re-cloned into a different vector (modified pET-49b) resulting in GST-only tagged proteins. This new vector was especially designed in-house by Vangelis Christodoulou and is simply called pGST. Once all constructs were successfully cloned into the pGST vector, the same small scale expression and purification screen was repeated. SDS-PAGE analysis of the induced cell samples and the elution can be seen in Figure 3-5. In this case, the expression level is extremely poor and basically not observable on the gels. Similar to the screen before, the gel of the elution sample shows still weak bands. The reason for the lack of expression is not known and possibly an error in the backbone of the new vector. If co-expressions needed to be performed the CARD proteins clearly would need to be re-cloned into a different vector to gain back the expression level as achieved with the pET49b vector.

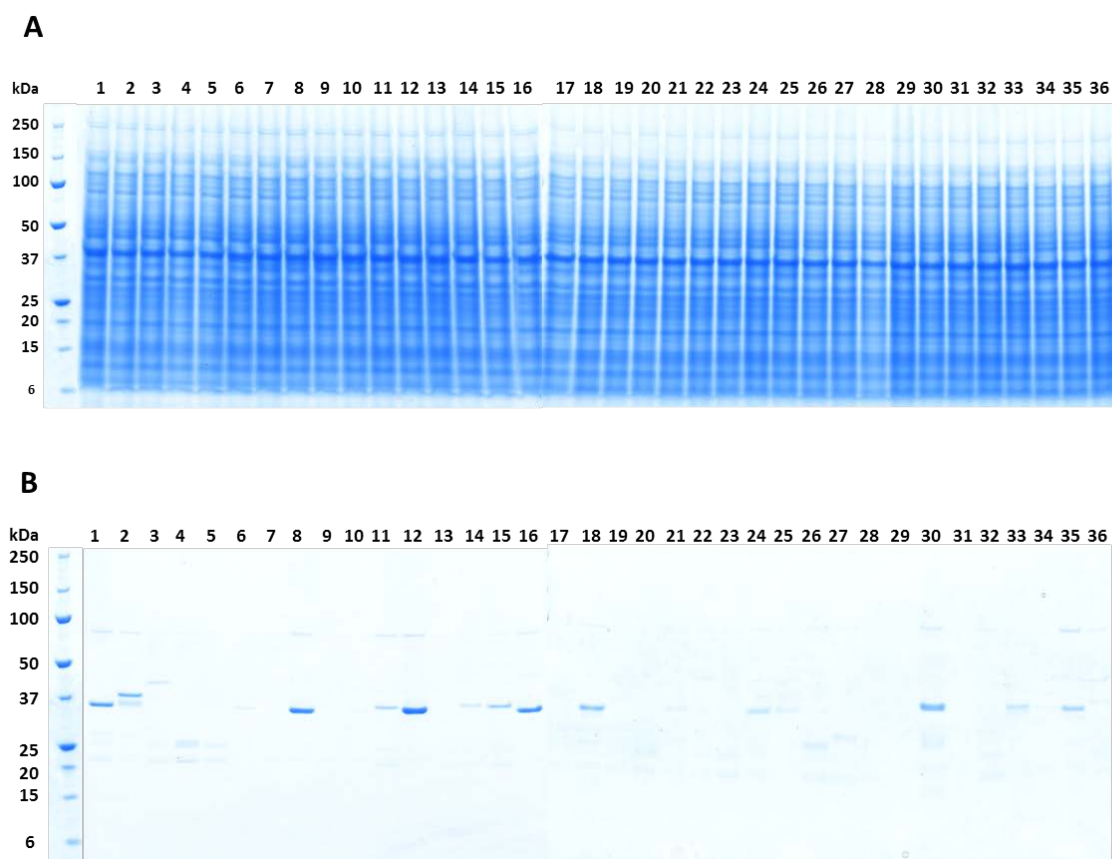


Figure 3-5 Small scale expression and solubility screen of GST-tagged CARD constructs using the new vector pGST. SDS-PAGE analysis of the samples collected from the induced cells (A) and from the eluted proteins from the G-sepharose columns (B). The constructs are numbered as listed in Table 3-2.

Taking a look at which His₆-CARDs turned out to be soluble, it was not surprising that most of them were CARDs whose structure has already been solved such as APAF1, ICEBERG, RAIDD and Rig-I. However, the screen also showed that other CARDs including CARD6, CARD9 and especially CARMA1 could be suitable proteins for further structural and biophysical studies. These are proteins for which no biophysical information was published yet. Unfortunately however, interaction partners of the CARDs (as stated in the literature), which show the highest solubility are in most cases insoluble or only a small fraction is soluble as listed in Table 3.3.

CARDs	CARD interaction partners and their solubility
NOD1 ✓	RIP2 ✗, Caspase9 ✓
ASC ✓	Caspase1 ✗, NALP1 ✓✓✓
APAF1 ✓✓✓	Caspase9 ✓
NALP1 ✓✓✓	Caspase9✓/APAF1 ✓✓✓, Caspase1 ✗
Caspase 9 ✓	APAF1 ✓✓✓
CARD8✓	Caspase1 ✗, Iceberg✓✓, COP1 ✗
CARD9 ✓✓	BCL10 ✗
RAIDD ✓✓✓	Caspase2 ✗
Iceberg ✓✓	Caspase1 ✗, CARD8 ✓
RIG-I ab ✓✓	MAVS ✗
MDA5 b ✓	MAVS ✗
CARMA1 ✓✓✓	BCL10 ✗
CARD6 ✓	-
ARC ✓	Caspase2 ✗

Table 3-3 Interaction partners of CARDs. Listed are the His₆-CARD constructs which were either soluble or only partially soluble together with their potential interaction partner(s). Results are illustrated as heavily ✓✓✓, moderately ✓✓ and minimally ✓ soluble constructs based on the visual analysis of the SDS-PAGE gels. Unexpressed/insoluble constructs are marked with ✗.

3.2.3 Co-expression and co-purification

The small scale expression and solubility screens showed that CARMA1- and CARD9-CARD could be suitable proteins for further investigation. However, both of these CARDS interact with BCL10-CARD, which could not be successfully purified in either screen. Therefore, co-expression was performed using BCL10 together with a selection of other CARDS in the hope that solubility could be increased by expressing BCL10 with its interaction partners CARMA1, CARMA2, CARMA3 and CARD9. Additionally, a number of other CARDS were selected (both soluble [APAF1] and insoluble constructs [e.g. Caspase-4]) for co-expression with BCL10 as negative controls. The CARD co-expression pairs were purified twice, once *via* GST-affinity chromatography and once *via* metal affinity chromatography. The SDS-PAGE analysis of the elution samples is shown in Figure 3-6 with the corresponding list shown in Table 3-4. Neither of the co-expressed CARDS increased the solubility of BCL10 significantly, nor could any BCL10 be pulled down with GST-CARMA1.

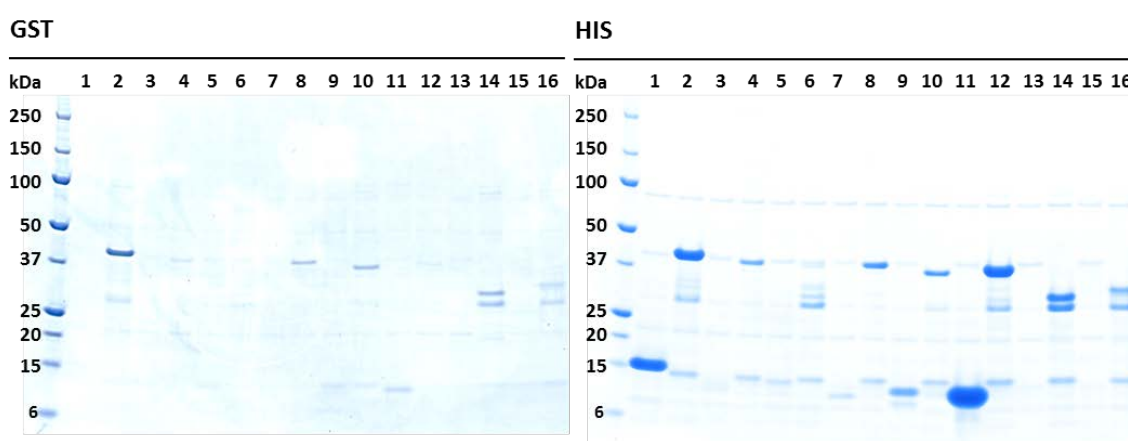


Figure 3-6 Small scale co-expression and co-purification. SDS-PAGE analysis of the elution samples collected from purifying the co-expressed proteins via G-sepharose beads (left) and Ni-NTA (right). The constructs are numbered as listed in Table 3-4.

#	Expressed		CARD used for purification	
	GST	His	Elution G-sepharose	Elution Ni-NTA
1	BCL10	CARMA1	-	CARMA1
2	CARMA1	BCL10	CARMA1	CARMA1
3	BCL10	CARMA2	-	-
4	CARD8	BCL10	(CARD8)	CARD8 / (BCL10)
5	BCL10	CARMA3	-	-
6	CARMA3	BCL10	-	-
7	BCL10	RIP2	-	-
8	CARD9	BCL10	CARD9	CARD9 / (BCL10)
9	BCL10	CARD6	-	CARD6
10	MAVS	BCL10	MAVS	MAVS / (BCL10)
11	BCL10	APAF1	-	APAF1
12	APAF1	BCL10	-	APAF1 / (BCL10)
13	BCL10	CASP4	-	-
14	CASP4	BCL10	-	-
15	BCL10	CASP5	-	-
16	CASP5	BCL10	-	CASP5 / (BCL10)

Table 3-4 Small scale co-expression and purification. BCL10 was co-expressed and purified with a number of selected CARDS in an attempt to increase solubility. Listed are the chosen pairs of CARDS as well as the constructs in the elution sample after the purification using either G-sepharose or Ni-NTA.

3.2.4 Large scale test expression and purification of selected CARDS

After the expression and solubility screens of the individual CARDS, several constructs were identified to be soluble. In order to obtain sufficient quantities of protein for further studies it was necessary to express and purify these constructs in a larger scale. The small scale purification protocol differs from standard protocols (e.g use of detergents or lysis method, further details can be as seen in section 2.3.6 and 2.3.7). To determine how well the small scale expression translates to larger scales, two GST-tagged CARDS were picked which are expected to behave differently: CARMA1, our best expressed CARD of the small scale screens and Iceberg, which showed

significantly lower levels of expression. GST-tagged CARMA1 and Iceberg were expressed on a 750 mL scale using BL21 (DE3) Gold cells with an induction temperature of 18 °C for 16 hours. After GST-affinity chromatography, the GST tag was removed by cleavage with 3C protease in batch overnight and the cleaved protein collected the next day. Individual steps of the purification are shown in Figure 3-7. CARMA1 could be eluted from the beads and was soluble at room temperature. In contrast, Iceberg was largely insoluble, similar to its behaviour in the small scale screen. It is important to point out that the sample of the beads with the bound GST-fusion protein can be misleading about the amount of soluble untagged protein, as fusion to GST, a 27.3 kDa protein, can increase solubility, whereas its removal may render the protein insoluble and lead to its precipitation on the beads. Large scale purification of His₆-tagged CARMA1, CARD9 and BCL10 has also been performed and is further described in Chapter 5.

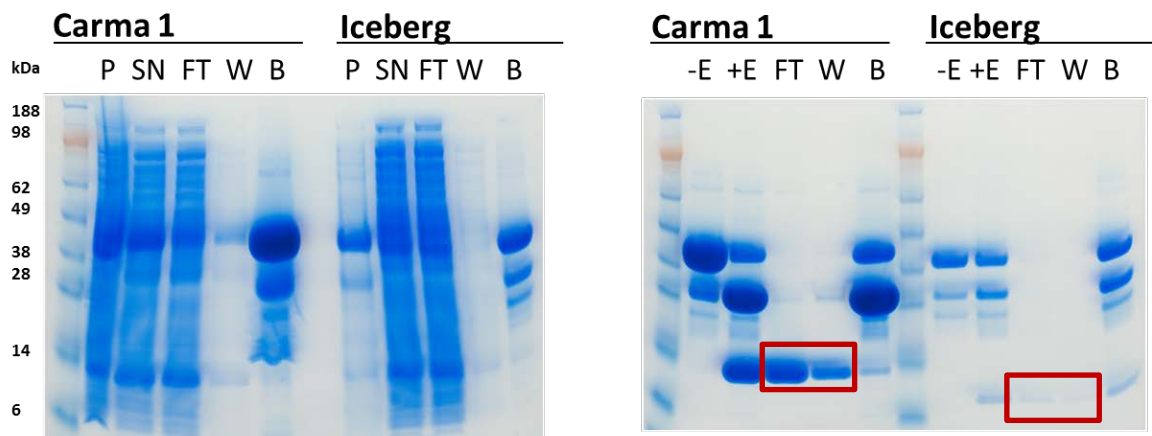


Figure 3-7 Purification of GST-CARMA1/-Iceberg. The purification steps were monitored by SDS-PAGE. P; pellet, SN; supernatant, FT; flow through, W; wash, B; beads, -E; batch sample before adding 3C-protease, +E; batch sample after adding 3C-protease, FT_E; eluate 1, W_E; eluate 2. The fractions collected containing the eluted proteins are marked with a red box.

In parallel with the GST-constructs, several His₆-CARDs were selected for large scale expression and purification by affinity chromatography. Proteins were eluted from Ni-NTA beads using concentrations of imidazole increasing up to 500 mM. SDS-PAGE analysis of the purification of CARD6, IPAF and ASC CARDS are shown in Figure 3-8. In all cases, the protein elutes during the wash step at a maximum concentration of 75 mM imidazole indicating the proteins may form soluble aggregates where His₆-tags are not fully exposed. The large scale purification of His₆-tagged CARMA1, CARD9 and BCL10 is further explained in Chapter 5.

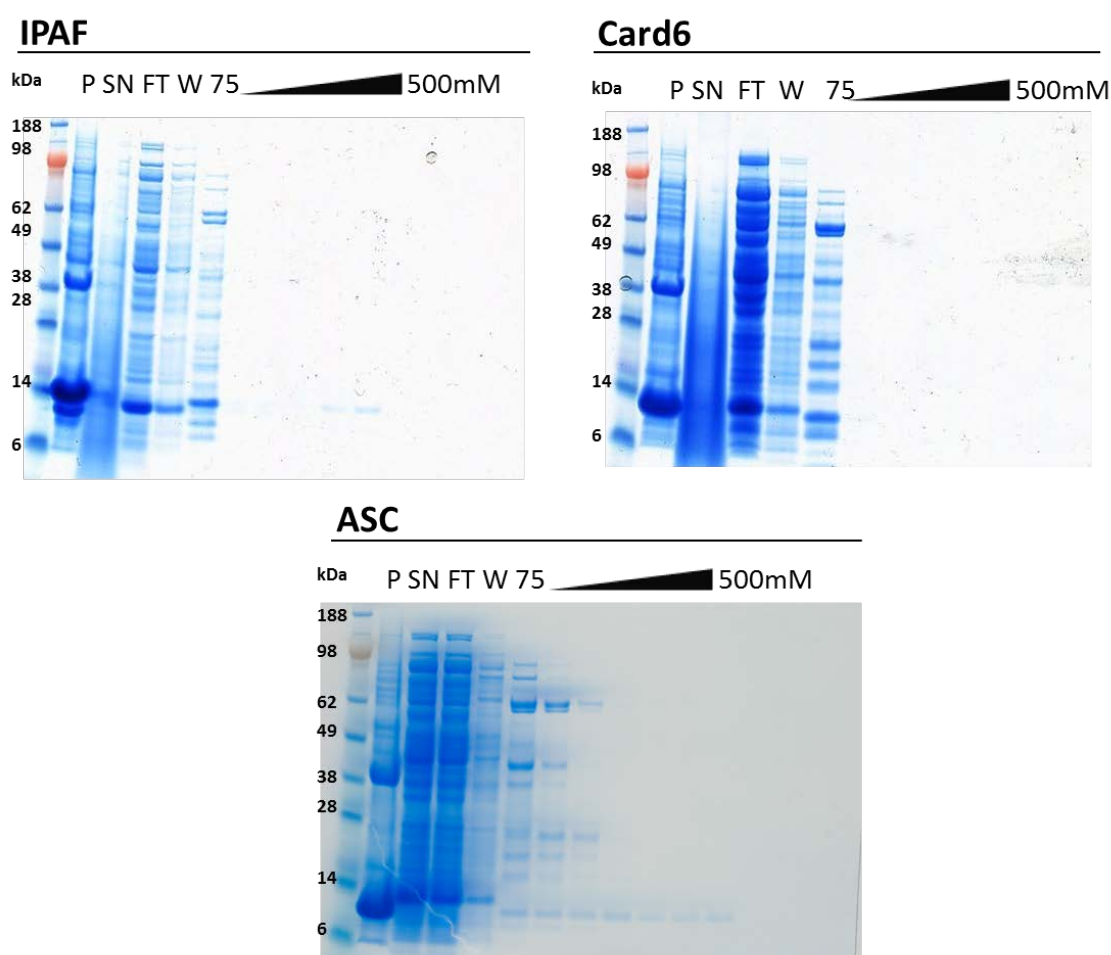


Figure 3-8 Purification of His₆ IPAF, -CARD6 and -ASC. The purification steps were monitored by SDS-PAGE. P; pellet, SN; supernatant, FT; flow through, W; wash, B; beads. The protein was eluted by an imidazole gradient from of 75 mM to 500 mM.

Regarding the disappointing expression level of the GST-tagged CARD constructs using the newly designed vector pGST, a complete re-cloning of all human CARDS into a new GST-vector for co-expression and pull-down screens was not taken any further because Qiao *et al.* (2013) [103] demonstrated the filament formation of BCL10-CARD at this point. This completely changed the aspect of how to approach the investigation of CARDS if the idea of filamentous CARD assembly wanted to be addressed as well. For this reason, the attempt to perform a wide range of co-expressions and pull-down experiments was dismissed and the focus was set on a selection of CARDS: Firstly, various experiments with CARMA1, CARD9 and BCL10 has already been performed (see Chapter 5) and provided a good system to look further into the polymerisation of CARDS. Secondly, another focus was set on the tandem CARDS of NOD2 as preliminary work has already been performed in our group and especially because the structure of the N-terminal NOD-CARDa had recently been solved by a former group member. This provided the opportunity to compare intra- and intermolecular CARD-CARD interactions (see Chapter 4).

3.3 Summary

CARDs are essential domains for mediating the formation of multi-protein complexes during intracellular signal transduction in innate and acquired immune signalling cascades. Therefore, it is of significant interest to understand the specificity and molecular features underlying these interactions. This project is focused on the biochemical and biophysical investigation of homotypic CARD-CARD interactions.

To gain insight into general properties of CARD-mediated protein interactions, all 33 human CARDs were cloned as His₆- and His₆-GST fusion proteins to perform co-expression and pull-down experiments *via* GST-affinity chromatography in order to identify soluble and interacting CARD pairs. Small scale expression and solubility screens of the individual domains, however, demonstrated a poor outcome in identifying soluble GST-constructs. To enable pull-down experiments via metal-affinity chromatography instead, all CARDs were re-cloned into a newly designed GST(-only)-vector. However, expression levels using this new vector were extremely poor.

On the other hand, solubility screens of the individual CARDs identified 14 His₆-constructs and 6 His₆-GST-constructs as potential candidates for expression in large-scale. Surprisingly, more His₆-constructs than GST-constructs turned out to be soluble or partially soluble. Our attention especially fell on the highly soluble CARMA1-CARD. More regarding this construct is described in Chapter 5.

Co-expression and pull-down experiments have not been taken any further at this point, as the EM analysis of BCL10-CARD filament lead us to the decision to focus on specific CARDs (NOD2, RIP2, CARMA1, CARD9, BCL10) to address this new aspect of complex formation *via* filamentous assemblies.

Chapter 4

Intra- and intermolecular CARD-CARD interactions of NOD2

4.1 Overview

NOD2, an intracellular pattern recognition receptor and member of the NOD-like-receptor family (NLR), is stimulated by peptidoglycans of the bacterial cell wall and believed to undergo ligand-induced oligomerisation through its NACHT domain, followed by the recruitment of RIP2 through CARD-CARD interactions, which ultimately leads to NF-kappaB activation [69]. Following up on previous work carried out in our group [131], which discovered an intramolecular interaction between the tandem CARDS of NOD2, the aim of my project was to biophysically and structurally investigate the interaction between these two domains.

Members of the death domain superfamily have been shown to engage in three distinct types of interactions to mediate formation of higher order complexes called type 1, 2 and 3 [117]. My goal was to investigate whether the intramolecular interaction between CARDa and CARDb of NOD2 is a type 1, 2 or 3 interaction or may be formed via a new binding interface. In addition to canonical 1:1 interactions as observed in the APAF1-Caspase-9 complex, members of the death domain superfamily have also been shown to be responsible for the formation of higher-order complexes such as the Apoptosome, MyDDosome, or even filaments [17, 54, 103]. Therefore, further questions were addressed: Are both NOD2-CARDS required to recruit the downstream effector RIP2-CARD? Does the CARD of RIP2 form higher order complexes and if so, does NOD2 enhance polymerisation?

The structure of NOD2-CARDa has recently been determined in our group. Based on this new structure, mutants have been generated in order to investigate the above raised questions using a range of biophysical techniques including ITC, NMR, CD as well as fluorescence spectroscopy.

4.2 The intramolecular interaction of the tandem CARDs of NOD2

As introduced in section 1.6.5.1, NMR data suggested that two NOD2-CARDa tryptophans (most likely W59, W63 or W68) may play a crucial role in binding of CARDb. To validate this observation, experiments needed to be designed to test if the intramolecular interaction between CARDa and CARDb may use binding surfaces either equivalent to type 1, 2, and 3 or equivalent to tandem DEDs.

4.2.1 NOD2 constructs

4.2.1.1 CARDa

All five possible tryptophan mutants were generated by replacing the residue with an alanine. Furthermore, in order to select CARDa residues that may be involved in type 1, 2, or 3 binding interfaces, structures of those death domain family members forming higher-order complexes were chosen and superimposed onto the NOD2-CARDa structure. The filamentous complex of MAVS-CARD was the first higher order CARD structure solved via x-ray crystallography. [133] Residues in each binding site of MAVS-CARD were highlighted before an overlay with CARDa was performed using the Dali pairwise alignment [148], with results evaluated in PyMOL [149] (Figure 4-1). Additionally, the same overlap was performed with other death domain superfamily complex structures such as Procaspase-9-APAF1 (type 1, APAF1 (PDB: 4RHW, chain A), Procaspase-9 (PDB: 4RHW, chain E)) or the MyDDosome (type 1,2 and 3, MyD88 (PDB: 3MOP, chain A), IRAK4 (PDB: 3MOP, chain G)) to select CARDa residues for mutational studies. Those residues potentially in the right place of a type 1, 2 or 3 interaction can be found in Table 4-1 with the sequence alignment of the template molecules shown in Figure 4-2. Interestingly, residues determined using the death domains of MyD88

and IRAK4 led to similar results as the overlap of CARDa with MAVS-CARD, even though these molecules belong to different subfamilies. The mutants E69K, E72K and R86A, which were also identified have already been generated and previously analysed in the group [131].

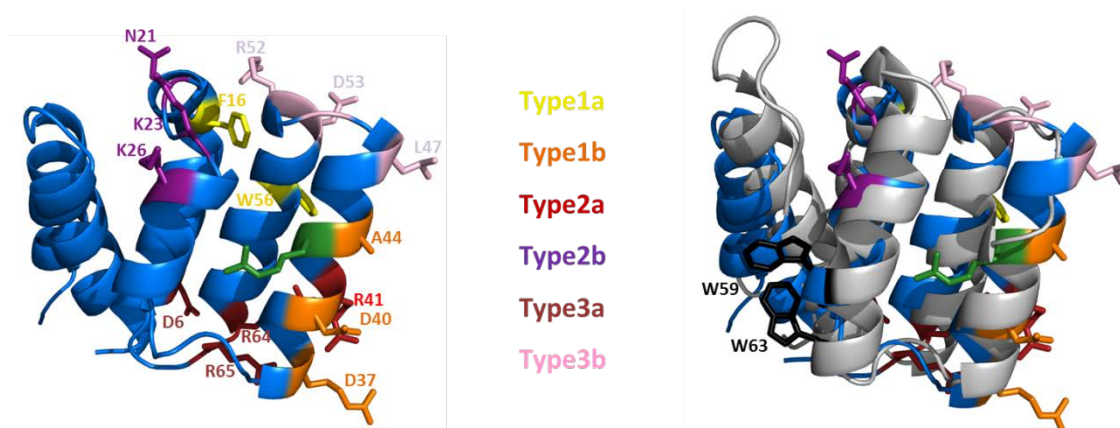


Figure 4-1 Residues of MAVS-CARD essential for type 1, 2 and 3 interactions. Left: To demonstrate the most important residues involved in type 1, 2 or 3 interaction of CARDa, we selected MAVS-CARD (PDB: 3J6J) and highlighted the residues of each binding site. Right: NOD2-CARDa in grey is superimposed with MAVS-CARD. The tryptophans W59 and W63 are highlighted in black. Figures were created using PyMOL.

	IRAK	MyD88	MAVS
Type1a	R38	R38	R86
Type1b	E64/ E69 / E72	E54 / D58 / Y71 / H75	Y71 / E72 / H75
Type2a	D90 / W93 / N94	N94 / K95	E69 / N94 / K95
Type2b	S49 / E51 / D113	S49 / D113	S47 / S49 / E51
Type3a	E69 / E72/ Q79	Q79	E69 / E51 / Q79
Type3b	S49 / Q79	-	S49 / Y71 / Q79

Table 4-1 Selected CARDa residues. Residues of CARDa which are potentially involved in a type 1,2 or 3 interaction, determined by superimposing CARDa with the death domains of IRAK4, MyD88 and the CARD of MAVS via Dali pairwise alignment and inspection in PyMOL. E69K, E72K and R86A have been already cloned and analysed by Fridh and Rittinger (2012).

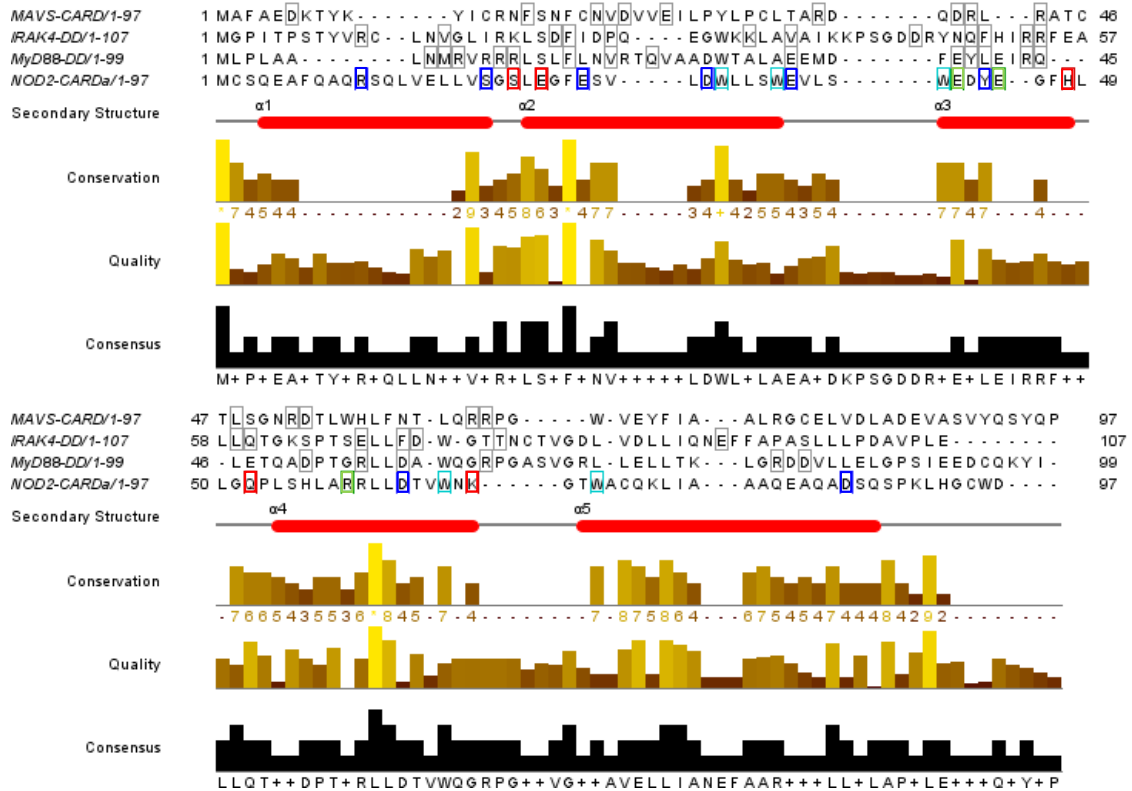


Figure 4-2 Design of NOD2-CARDa mutants. The design of CARDa mutants was based on an overlap with available structures of complexes formed by the death domain superfamily; MyD88 (PDB: 3MOP, chain A), IRAK4 (PDB: 3MOP, chain G) and MAVS (3J6J). Residues chosen for site-directed mutagenesis are in red, whereas the green residues have already been tested by Dr Veronica Fridh [131]. The tryptophans of CARDa are shown in cyan. All overlays have been performed using the Dali pairwise alignment. Figure was completed using JalView [150], with primary sequence alignment via Muscle [151]. The secondary structure was added manually to this figure and corresponds to the structure of NOD2-CARDa.

After performing site-directed mutagenesis; a small scale expression and solubility screen was performed to determine which mutants are suitable for further experiments (SDS-PAGE analysis shown in Appendix A 4-1). A list of all mutants tested and their solubility can be found in Table 4-2. Mutants selected for further investigations were: W59A, W63A, W98A, S49A, H75A, Q79A and N94A.

Mutation	Solubility	Potential interaction site
S49A	✓	Type2b/3b
N94A	✓✓	Type2a
Q79A	✓	Type3a/3b
Q79K	x	Type3a/3b
H75A	✓✓	Type 1b
H75Q	✓✓	Type 1b
K95A	x	Type 2a
E51A	x	Type 2b/3a
E51K	x	Type 2b/3a
W59A	✓✓	Tryptophan
W63A	✓	Tryptophan
W68A	x	Tryptophan
W93A	x	Tryptophan
W98A	✓✓✓	Tryptophan

Table 4-2 NOD2-CARDab mutants with mutations in CARDa. Mutations listed are all located in CARDa and have been analysed by SDS-PAGE for solubility. Results are illustrated as heavily ✓✓✓, moderately ✓✓ and minimally ✓ soluble constructs based on the visual analysis of SDS-PAGE gels. Insoluble constructs are marked with x.

4.2.1.2 CARDb

From previous work, it was known that CARDb alone is an unstable protein, which easily precipitates. Therefore, any mutations inserted in CARDb of the tandem CARD constructs were likely to have a severe effect on its solubility and stability. However, given the lack of information regarding the binding surface on CARDb, it was worth trying to generate mutants, which may be sufficiently soluble and stable for further experiments. In the absence of a structure of NOD2 CARDb, a homology model was generated in SwissModel in order to select surface residues for mutagenesis [152]. The solution structure of BinCARD (PDB: 4DWN) was found to have the highest z-score with NOD2-CARDa as a result of a Dali structural alignment [148] and was therefore used as template.

To predict the binding site of CARDb and to be able to design appropriate CARDb mutant constructs, CARDa was aligned with available structures of tandem CARDb mutant constructs, CARDa was aligned with available structures of tandem domains. In these alignments, the focus was set on the its tryptophans W59 and W63: When NOD2-CARDa was aligned with RIG-I tandem CARDs (PDB: 4A2Q), the tryptophans were not located near the binding site of RIG-I-CARDa and b. Interestingly, further alignments have revealed that when NOD2-CARDa is aligned with the tandem DEDs of MC159 (PDB: 2BBZ), residue W59 and W63 are in the position of the DED-interface as shown in Figure 4-3. DED1 is highly divergent from a typical DED fold; therefore, the interaction differs from the known modes of interaction within the DD superfamily. Both CARDa and the model CARDb were then superimposed with the MC159 tandem DEDs to select residues of CARDb, which are potentially at the interface (Figure 4-4). D133, H137, R138 and R182 have been chosen for site-directed

mutagenesis, whereas E158 and Q201 were added as negative controls, located at the opposite end of the molecule.

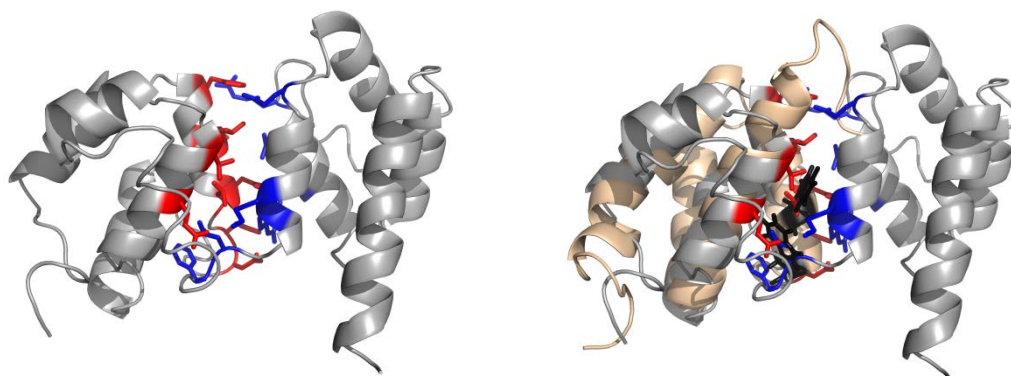


Figure 4-3 MC159 tandem DED. Left: The tandem DED of MC159 (grey, PDB: 2BBZ) with the interface highlighted in red and blue. Right: NOD2-CARDa in gold superimposed with tandem DEDs of MC159. Tryptophans W59 and W63 of NOD2-CARDa are highlighted in black. Figures were generated using PyMOL.

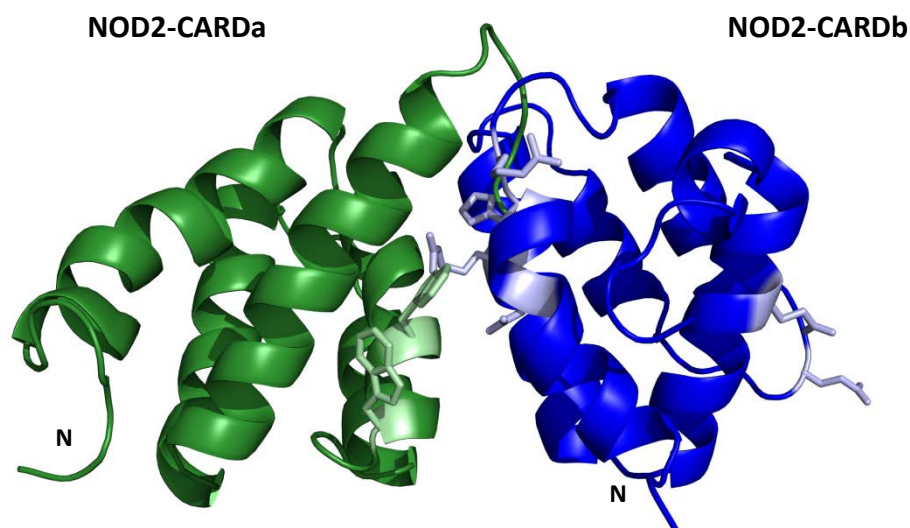


Figure 4-4 Model of the tandem CARDs of NOD2. The recently determined solution structure of NOD2-CARDa (green) in combination with a model of NOD2-CARDb (blue, created with SwissModel). The orientation between the two domains is based on the tandem DEDs of MC159. Highlighted are the positions of tryptophan W59 and W63 in light green, which are

believed to play a crucial role in CARDb binding. Residues in CARDb chosen for site directed mutagenesis are shown in light blue. Figure was generated using PyMOL.

After performing site-directed mutagenesis to insert mutations in CARDb of the NOD2 tandem CARD constructs, a small scale expression and solubility screen was performed to determine which mutants to focus on (SDS-PAGE analysis data shown in Appendix A 4-2). A list of the mutants produced and their solubility can be found in Table 4-3. Mutants selected for further investigations were: D133A, H137A, E158A and Q201A.

Mutation	Solubility
D133A	✓
H137A	✓
H137Q	✓
R138A	✗
R182A	✗
E158A	✓
E158K	✗
Q201A	✓✓
Q201K	✓✓✓

Table 4-3 NOD2-CARDab mutants with mutations in CARDb. Mutations listed are located in CARDb and have been screened regarding their expression and solubility. Solubility is based on the visual analysis of SDS-PAGE gels.

4.2.2 NMR titration experiments with the tryptophan CARDa mutants

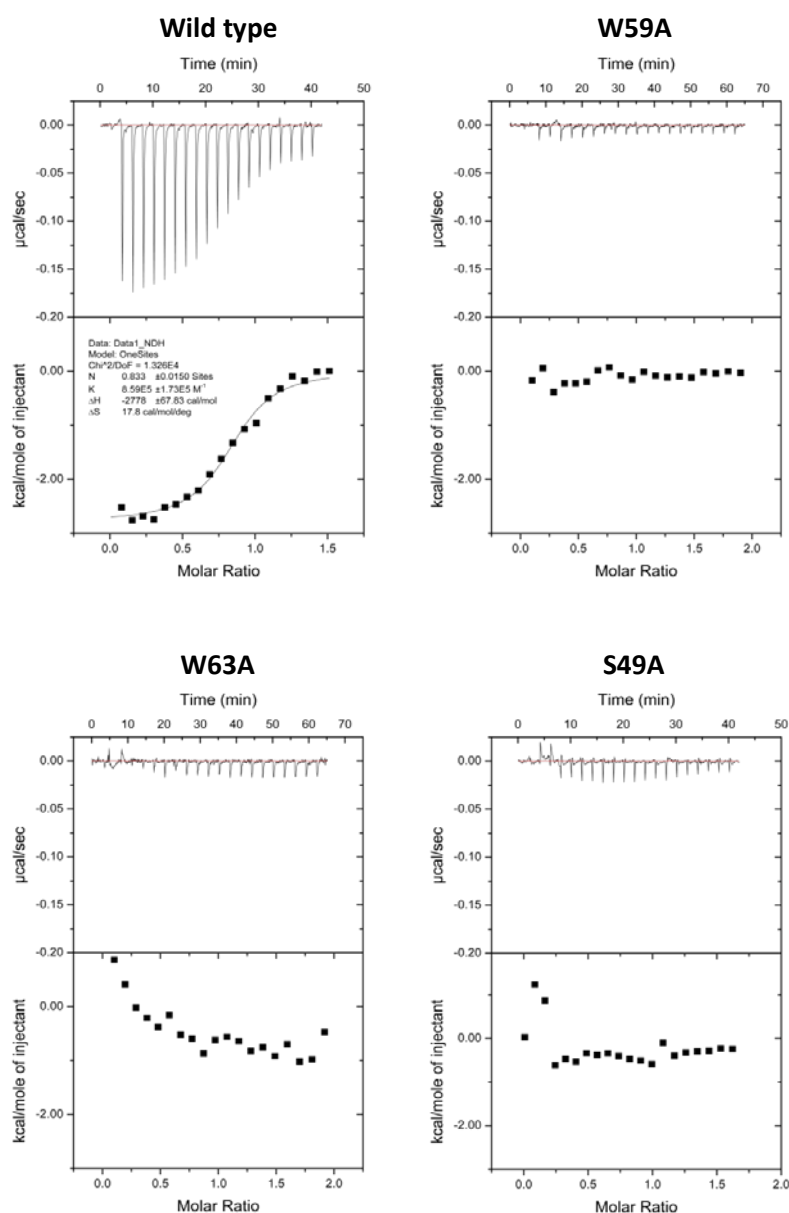
NMR is a very powerful technique to map residues of the interaction site of a protein upon complex formation with its binding partner. The experiments recorded in this case are ^{15}N -HSQCs. The chemical shifts of the N-H cross correlations are sensitive to their chemical environment and hence titrating the binding partner to the ^{15}N -labelled sample will change the environment of the residues involved in complex formation. This can be monitored as chemical shift changes in both dimensions when the spectra before and after the addition of the ligand are superimposed [153]. The NMR titration

experiments with the individual NOD2 CARDS were first performed by Kovilen Sawmynaden, a former group member. Spectra of a ^{15}N -NOD2-CARDa sample were recorded before and after unlabelled NOD2-CARDb was added. Protein samples were in 100 mM NaCl, 20 mM Na-Phosphate pH 7.1 and 2 mM TCEP and experiments were recorded at 25 °C. Full spectra of these experiments can be seen in Appendix A 1-1. When CARDb was titrated into CARDa, the majority of the peaks are changing their chemical shifts and therefore it is not possible to use the assigned CARDa spectrum to identify single peaks of the complex spectrum. However, the saturation transfer difference experiment shown in Figure 1-18, indicated that the tryptophans are playing a key role in CARDb binding. The chemical shift changes of two tryptophans are clearly visible when the spectra of 'CARDa' and 'CARDa + CARDb' are superimposed (Figure 1-19). Based on these observations, my goal was to repeat the titration experiments using different tryptophan mutants. Examination of mutants lacking a specific tryptophan should give insight into which of the two tryptophans is at the CARDa-CARDb interface. However, recombinant expression of the tandem CARDS (to obtain the single CARDS) was significantly reduced in minimal (M9) media. A commercial 'rich' isotopically-labelled medium was sourced to see if this could alleviate the problem. However, this did not appear to solve the observed problem with protein expression and hence, the amount of bacterial culture required to obtain sufficient protein for NMR experiments became prohibitively expensive. Using a maximum of 6 litres per tryptophan mutant, it became clear that the mutants' expression was still too low and the protein yield (e.g end concentration of the sample NOD2-CARDa W63A: 18 μM in 250 μL) not sufficient to complete the titration experiments as the signal to noise ratio was too poor.

4.2.3 Isothermal titration calorimetry with CARDa mutants

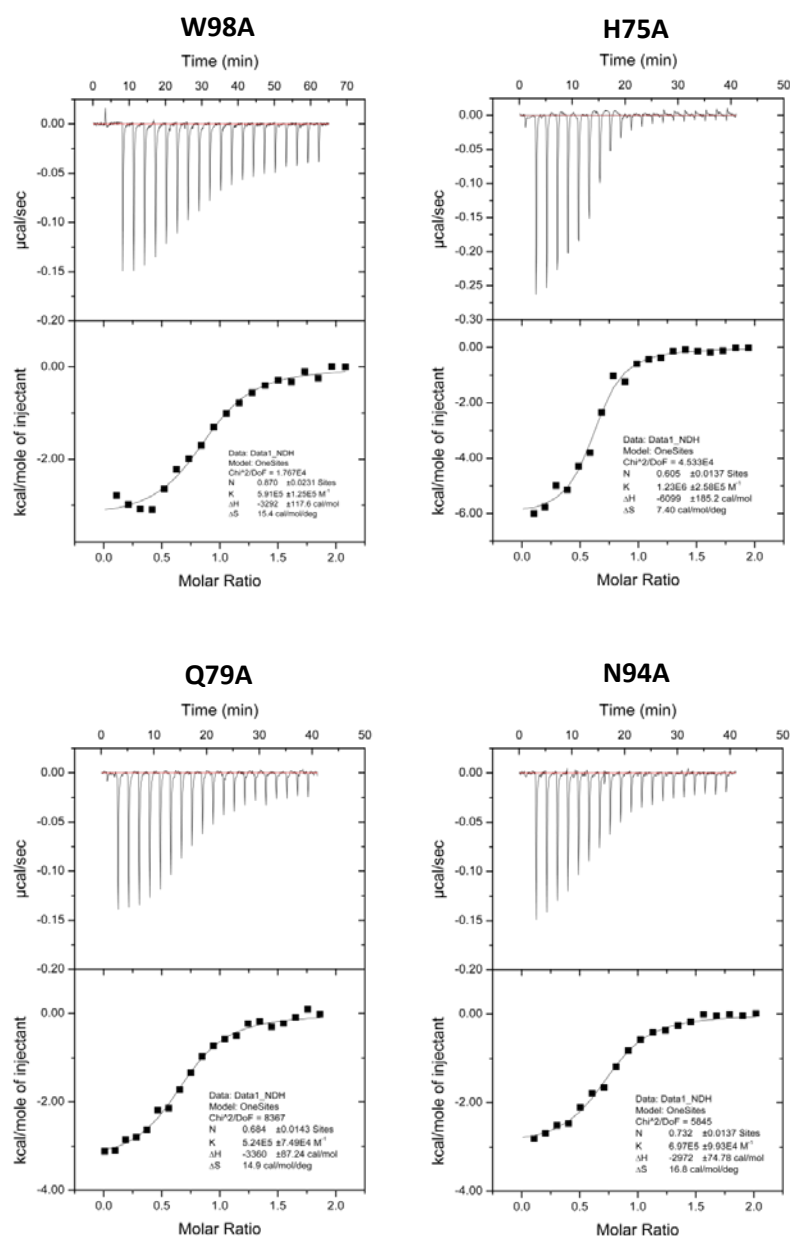
Due to the problems associated with the expression of ^{15}N -labelled NOD2 CARDS, it was decided to further investigate the NOD2 mutants by isothermal titration calorimetry (ITC) using unlabelled proteins. ITC is a technique, which can be used to measure the binding affinity and thermodynamic parameters of a protein-protein interaction. Titration of NOD2-CARDa WT ($>200\ \mu\text{M}$) into NOD2-CARDb WT at $25\ ^\circ\text{C}$ has previously been shown to be exothermic with a K_D of $1\ \mu\text{M}$ and a binding stoichiometry of 1:1 ($n = 0.8\text{-}1.0$) [131]. All of the following experiments have been performed using the same experimental conditions (50 mM Tris pH 7.5, 100 mM NaCl, 2 mM DTT) and were done in triplicate.

Fridh *et al.* (2012) already performed ITC experiments with NOD2-CARDa E69K, E72K and R86A. These residues were selected to interrupt a type 1 interaction. However, titration with wild type CARDb showed that these residues in CARDa are not important in the binding of CARDb. My goal was to test further mutants which were predicted to be involved in a type 2 and 3 interaction as well as the tryptophans mutants described above. Mutants tested were W59A, W63A, W98A, S49A, H75A, Q79A and N94A. The individual titrations are presented in Figures 4-5 and 4-6, including the calculated stoichiometry and thermodynamic parameters. Titrations performed with the wild type, H75A, Q79A, N94A and W98A resulted in exothermic reactions of complex formation. Binding constants and thermodynamic parameters derived from titrations were calculated after fitted to a single-site binding model, with $K_D = 1/K_A$ and $\Delta G = -RT\ln K_A$. Standard deviations were calculated between triplicate measurements. Dissociation constants varied between 0.6 and $2.3\ \mu\text{M}$.



	N	K _A [M ⁻¹]	ΔH [cal mol ⁻¹]	ΔS [cal mol ⁻¹ K ⁻¹]	K _D [μM]	ΔG [kcal mol ⁻¹]
WT	0.81 ± 0.004	9.60E+05 ± 2.8E+04	-3031 ± 10	17.2 ± 0.1	1.05 ± 0.03	-8.16 ± 0.02
W59A	-	-	-	-	-	-
W63A	-	-	-	-	-	-
W49A	-	-	-	-	-	-

Figure 4-5 ITC analysis of NOD2-CARDa mutants. The upper plot shows the raw data for the titration of NOD2 CARDa into CARDb at 25 °C. The lower plot shows the binding isotherm for the interaction. Just one experiment per mutant is shown above, however, listed are the binding constants and thermodynamic parameters derived from three titrations by calculating the average. Data were fitted to a single-site binding model. Standard deviations were calculated between triplicate measurements



	N	K _A [M ⁻¹]	ΔH [cal mol ⁻¹]	ΔS [cal mol ⁻¹ K ⁻¹]	K _D [μM]	ΔG [kcal mol ⁻¹]
W98A	0.86 ± 0.007	6.25E+05 ± 4.7E+04	-3181 ± 35	15.8 ± 0.2	1.61 ± 0.12	-7.90 ± 0.05
H75A	0.66 ± 0.004	1.57E+06 ± 4.5E+04	-6000 ± 62	8.2 ± 0.2	0.65 ± 0.02	-8.45 ± 0.02
Q79A	0.67 ± 0.006	4.54E+05 ± 3.5E+04	-3365 ± 38	14.6 ± 0.2	2.27 ± 0.17	-7.71 ± 0.05
N94A	0.75 ± 0.004	7.21E+05 ± 2.1E+04	-3045 ± 32	16.6 ± 0.1	1.46 ± 0.04	-7.99 ± 0.02

Figure 4-6 ITC analysis of additional NOD2-CARDA mutants. The upper plot shows the raw data for the titration of NOD2 CARDA into CARDB at 25 °C. The lower plot shows the binding isotherm for the interaction. Just one experiment per mutant is shown above, however, listed are the binding constants and thermodynamic parameters derived from three titrations by calculating the average. Data were fitted to a single-site binding model. Standard deviations were calculated between triplicate measurements

Due to the difficulties with the stability of CARDb the stoichiometry of complex formation n was < 0.7 in most experiments. A better value of $n = \sim 0.9$ could be achieved with the assumption that the cell sample concentration of CARDb was lower than the original value (with a proportion of CARDb partially misfolded). The heat generated with NOD2-CARDa mutants W59A, W63A and S49A was not sufficient to allow fitting of the data. No improvement in signal could be achieved when the experiments were repeated at different temperatures (12 °C and 18 °C). This indicates that the mutations W59A, W63A and S49A interrupt complex formation. These mutants have been structurally analysed via circular dichroism (far-UV spectra) and NMR (1D proton spectra) to confirm that these proteins are all folded (data not shown).

4.2.4 Circular Dichroism to identify interacting CARDb residues

Circular dichroism (CD) was previously used to analyse the fold and stability of the individual NOD2 CARDS as well as the tandem construct [131]. Protein stability was assessed by thermal unfolding of the proteins between 5 and 95 °C with measurements at 222 nm (Figure 1-16). Protein concentration was 0.15 mg/ml in CD buffer (150 mM NaCl, 25 mM Tris-HCl pH 8). NOD2-CARDa unfolded with a midpoint (T_m) at 66.2 °C whereas CARDb unfolded with a midpoint at 38.0 °C. However, the tandem construct showed two transitions occurring at higher temperatures ($T_{m1} = 55.6$ °C and $T_{m2} = 80.8$ °C) than for the individual domains.

To determine the orientation of the tandem CARDS of NOD2 with respect to one another, it was important to gain information not only about CARDa residues involved in the interface, but also of CARDb. Up to this point, only wild type CARDb had

been purified. From previous work we knew that CARDb alone is an unstable protein, which shows precipitation after recording experiments at 25 °C. Therefore, it could be expected that a mutation interrupting the intramolecular interaction may lead to a misfolded CARDb, which in turn would precipitate upon cleavage from the stabilising CARDa. However, small scale solubility screen showed that the majority of the tandem CARD constructs with mutations in CARDb are soluble and we were able to purify the tandem CARD constructs with single mutations (located in CARDb) D133A, H137A, E158A and Q201A. Based on the ITC results with CARDa mutants, the tandem constructs containing wild type and W98A were chosen as a positive controls and W59A and S49A as negative controls. To our surprise, the obtained melting curve of the various tandem NOD2 constructs all showed a single transition point with midpoints all between 59 °C and 62 °C (Figure 4-7).

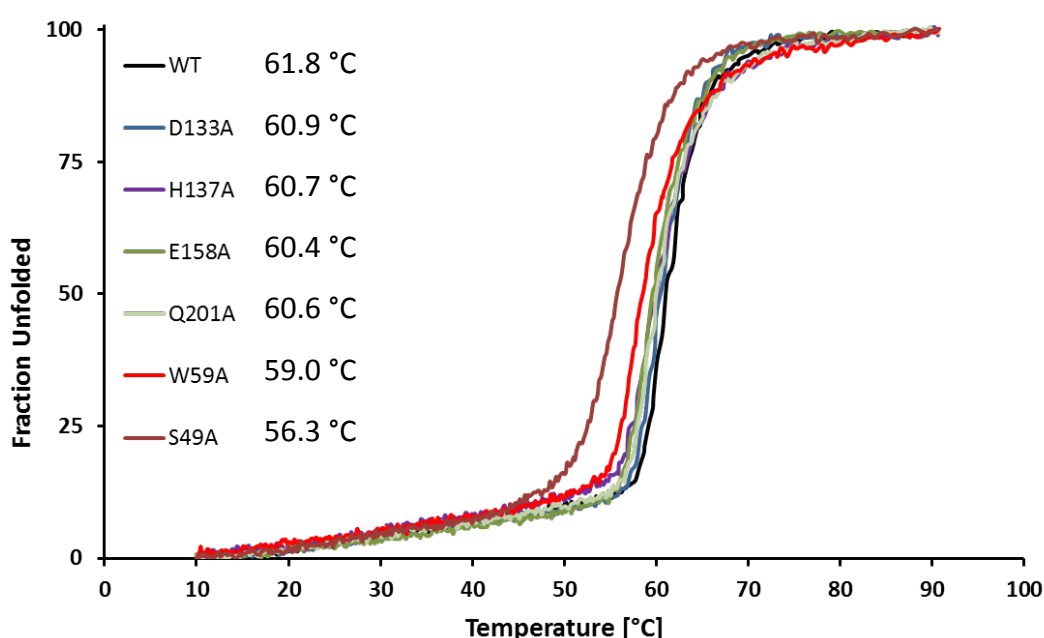


Figure 4-7 CD thermal unfolding of NOD2 tandem CARD mutants. Thermal unfolding experiments of various NOD2 tandem CARD mutants measured between 10 °C and 90 °C at 222nm showed single transition points within 59 °C and 62 °C.

The melting profile obtained differed from what we expected (Figure 1-16) and therefore, CD thermal unfolding experiments were repeated with the single domains of the wild type CARDa, CARDb and both CARDS mixed in at equimolar concentration prior the experiment. The resulting curves are presented in Figure 4-8. NOD2-CARDa unfolded with a midpoint (T_m) at 74.3 °C whereas CARDb unfolded with a midpoint at 43.5 °C. The tandem CARD construct showed one transition occurring at 61.6 °C. If both domains were purified separately and then mixed, the thermal stability was higher than CARDb alone although the melting point is 51.8 °C, which is a similar to the result obtained by Fridh *et al.*

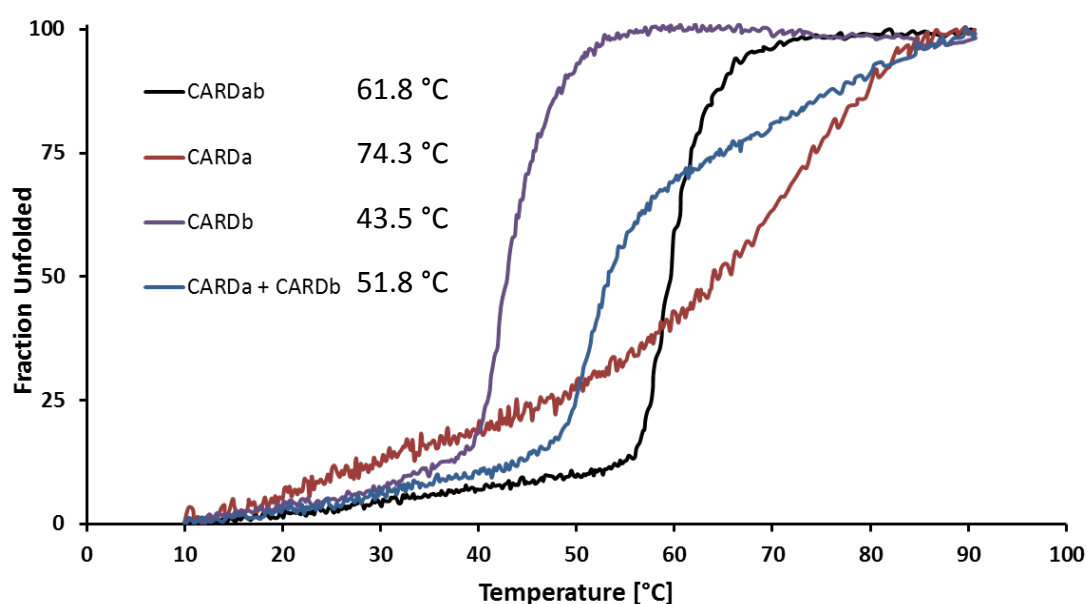


Figure 4-8 Thermal unfolding of wild type constructs. Thermal unfolding of the samples was performed at a single wavelength (222 nm) between 10 and 90 °C.

We can only speculate about the reasons for the discrepancy between the data obtained in this work and the data by Fridh *et al.*, as equivalent conditions and instrumentations were used. One possibility is that the protein sample used by Fridh *et al.* was contaminated. A contamination could have interfered with the thermal unfolding of the NOD2 tandem CARDS by stabilising CARDa, which then would increase the melting temperature resulting in two transition points.

Unfortunately, the difference of the various midpoints obtained here between the different mutants are not significant enough to draw any conclusions. A single mutation does not seem to interrupt the complex nor destabilise the protein significantly. This even applies to the W59A and S49A mutants, which were shown *via* ITC to be essential for the CARDb interaction.

The main difference between the CD experiments performed and the ITC experiments, is that the CARDS were still connected with the linker in the CD experiments whereas they were separate domains in the ITC titrations. For this reason, the CD experiments were repeated for several mutants, although this time including the thrombin site between CARDa and CARDb. This would allow purification of the domains together, before separating them by cleaving the linker with thrombin in hope that the interaction will be weakened and a difference between the mutants can be observed. However, the melting curves obtained for the tandem CARD mutants after thrombin cleavage result in midpoints that still only vary between ~3 °C, although overall they are lower and between 52 °C and 55 °C (Figure 4-9). The midpoints obtained in this experiment were similar to the values measured when CARDa and CARDb were individually purified and mixed prior the thermal unfolding experiment.

Based on these results it was concluded that CD is not providing significant differences between the mutants to draw any conclusions and suggesting that one mutation is not sufficient to disrupt the interaction enough to monitor stability differences *via* thermal unfolding.

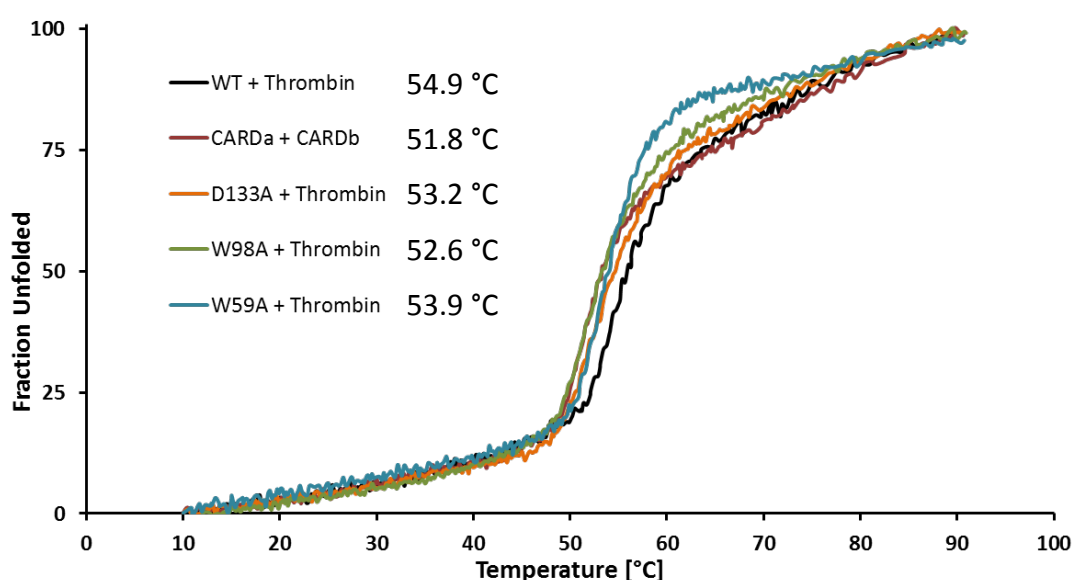


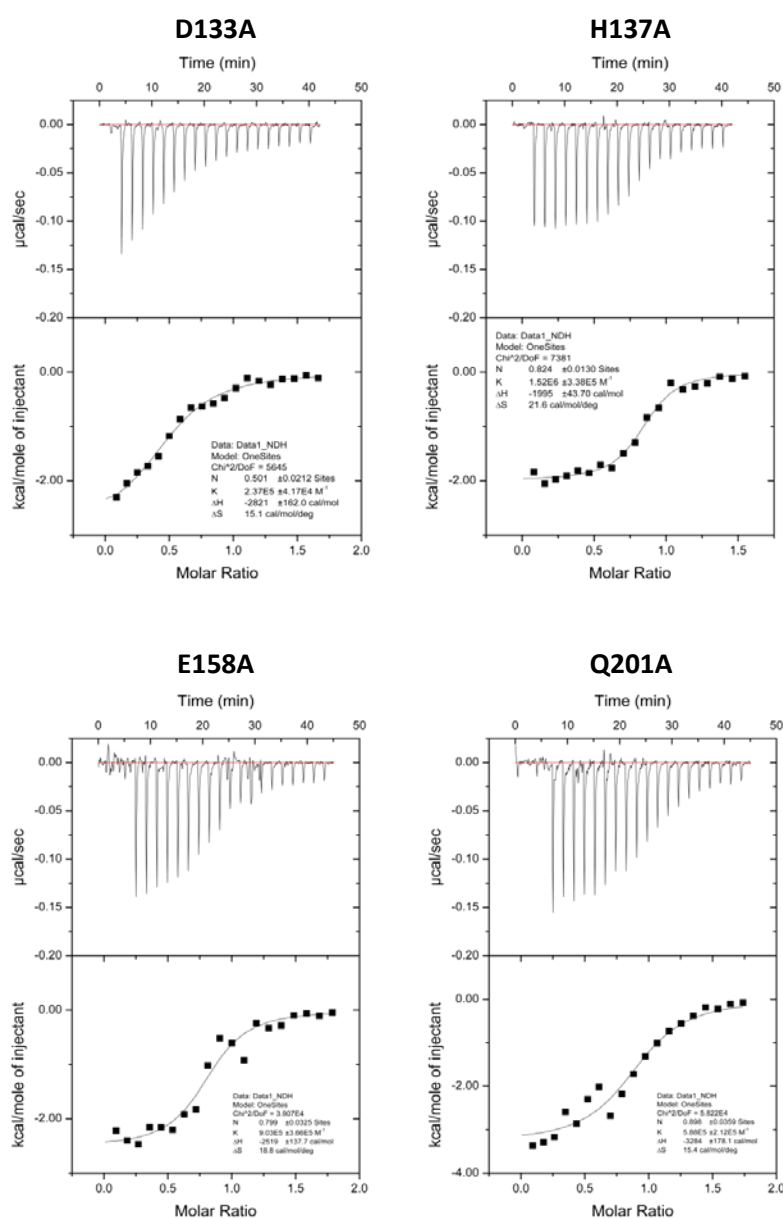
Figure 4-9 CD thermal unfolding of NOD2 tandem CARD mutants after thrombin cleavage.

The tandem NOD2-CARDab mutants contain a thrombin cleavage site between the two domains. Prior to CD thermal unfolding experiments, the sample was cleaved with thrombin. The thermal unfolding profile of the individually purified CARDa and CARDb mixed together is included to demonstrate the similarity of the curves and midpoints.

4.2.5 Isothermal titration calorimetry with CARDb mutants

Given that CD has not provided a clear answer if any NOD2-CARDb mutations are having an effect on CARDa binding, it was decided to test if the mutated CARDb domains can be cleaved from CARDa and further purified individually to perform ITC experiments as described with the CARDa mutants in section 4.3.3. Fortunately, it was

possible to complete ITC experiments using wildtype CARDa by titrating it into the sample cell containing the CARDb mutants D133A, H137A, E158A or Q201A. The individual titrations and the thermodynamic parameters are presented in Figure 4-10. Titrations performed with all four CARDb mutants resulted in an exothermic reaction of complex formation. The binding curves could be fitted to a single-site binding model with a dissociation constant of 0.9 - 3 μ M. As expected, some CARDb mutants did not seem very stable at room temperature. The noise level of the baseline was significantly higher than in previous experiments with CARDa mutants. The instability of D133A is represented in the stoichiometry of $n = 0.53 \pm 0.004$. Similar to the experiments with CARDa mutants, the assumption of a lower actual concentration of CARDb resulted in a better value of up to $n \sim 0.85$, implying that a part of the D133A cell sample was not properly folded and functional.



	N	K _A [M ⁻¹]	ΔH [cal mol ⁻¹]	ΔS [cal mol ⁻¹ K ⁻¹]	K _D [μM]	ΔG [kcal mol ⁻¹]
D133A	0.53 ± 0.004	4.05E+05 ± 7.9E+04	-2444 ± 38	17.3 ± 0.4	2.92 ± 0.5	-7.65 ± 0.12
H137A	0.79 ± 0.008	1.15E+06 ± 5.5E+04	-2100 ± 38	20.6 ± 0.2	0.95 ± 0.04	-8.27 ± 0.03
E158A	0.77 ± 0.011	7.51E+05 ± 1.4E+05	-2800 ± 95	17.3 ± 0.5	1.68 ± 0.25	-8.02 ± 0.11
Q201A	0.97 ± 0.010	1.60E+06 ± 5.4E+05	-2881 ± 71	18.3 ± 0.7	1.04 ± 0.24	-8.47 ± 0.21

Figure 4-10 ITC analysis of NOD2-CARDB mutants. The upper plot shows the raw data for the titration of NOD2 CARDa into CARDB at 25 °C. The lower plot shows the binding isotherm for the interaction. Just one experiment per mutant is shown above, however, listed are the binding constants and thermodynamic parameters derived from three titrations by calculating the average. Data were fitted to a single-site binding model. Standard deviations were calculated between triplicate measurements

4.3 Intermolecular CARD-CARD interactions of NOD2 and RIP2

Having demonstrated that CARDa residues W59A, W63A and S49A are probably involved in the intramolecular CARD-CARD interaction of NOD2 CARDa and CARDb, the next goal was to map residues involved in the binding site of the NOD2-RIP2 CARD-CARD intermolecular interaction. This would allow the generation of a model to gain insight into the architecture of the NOD2-RIP2 interaction. With recent literature suggesting that CARDS form higher order complexes [54], we were interested to test if both CARDS of NOD2 are essential for RIP2 recruitment. Additionally, highly insoluble CARDS such as BCL10 have recently been shown to form filaments by Qiao *et al.* (2013) [103], with filament formation significantly enhanced in the presence of the upstream scaffold protein CARMA1. With RIP2 having proved to be a problematic construct to work with and being a downstream effector molecule in the NOD2 signalling pathway, the question arose if RIP2-CARD also forms filaments and if they are nucleated by NOD2.

4.3.1 Pull-down experiments with NOD2 tandem CARD mutants

4.3.1.1 Constructs used

Previous work carried out in our group on the interaction between NOD2 and RIP2 involved co-expression and pull-down studies [131]. Based on results with the mutants tested so far, it was suggested that the tandem CARDS of NOD2 use the basic residues R38 and R86, located on the CARDa domain, for the interaction with RIP2 CARD. The majority of the new NOD2 tandem CARD mutants (amino acid (aa) 28-218, pGEX-6P1, GST-fusion) already created for section 4.3 has shown that the residues tested are not

involved in the NOD2 intramolecular interaction, therefore, these constructs were further analysed in this study for RIP2 binding.

Former group members have shown that RIP2-CARD cannot be expressed and purified without a solubility enhancement tag such as GB1 (Immunoglobulin binding domain of Streptococcal protein G) or MBP. In light of these observations RIP2-CARD aa 432-529 was cloned into pET-XM1 (kind gift of Sebastian Charbonnier, CNRS, Illkirch), an expression vector generating proteins with an N-terminal MBP-tag. MBP, which has a molecular weight of 40 kDa, is a highly soluble protein and is often used as a solubility enhancer of their fusion partner.

4.3.1.2 Small scale co-expression and pull-down experiments

The two vectors containing either GST-NOD2-CARDab or MBP-RIP2-CARD were co-transformed into BL21 (DE3) Gold cells and plasmid stability was maintained by addition of kanamycin and carbenicillin to the growth media of 150 mL LB. After induction and incubation at 18 °C overnight, the cells were harvested and lysed by sonication in 150 mM NaCl, 50 mM TRIS-HCl pH 8, 2mM DTT. The supernatant was loaded onto 150 µL GST-sepharose beads solution. Pull-down experiments were carried out using 5 mL polystyrene columns (Sigma Aldrich). Beads were washed with three times 1 mL lysis buffer. Elution was carried out by adding 10 mM reduced glutathione to the buffer. Samples for SDS-PAGE analysis were taken at every step and the pull-down experiment with wild-type GST-NOD2-CARDab/MBP-RIP2-CARD is shown in Figure 4-11 A. Expression analysis showed that both proteins were expressed although RIP2-CARD was overexpressed to a greater degree. The gel shows that MBP-RIP2-CARD clearly bound to the NOD2 protein. Control experiments were

performed to exclude any non-specific binding of the MBP fusion tag by using wild-type NOD2-CARDab co-expressed and purified with only MBP (Figure 4-11 B). The same pull-down experiments were repeated with all NOD2-CARDab mutants available and are shown in Figure 4-12. Mutant R38A, originally shown in Fridh *et al.* (2012) to lose the ability to bind RIP2, was used in this study as a negative control experiment. All new mutants were able to bind RIP2-CARD.

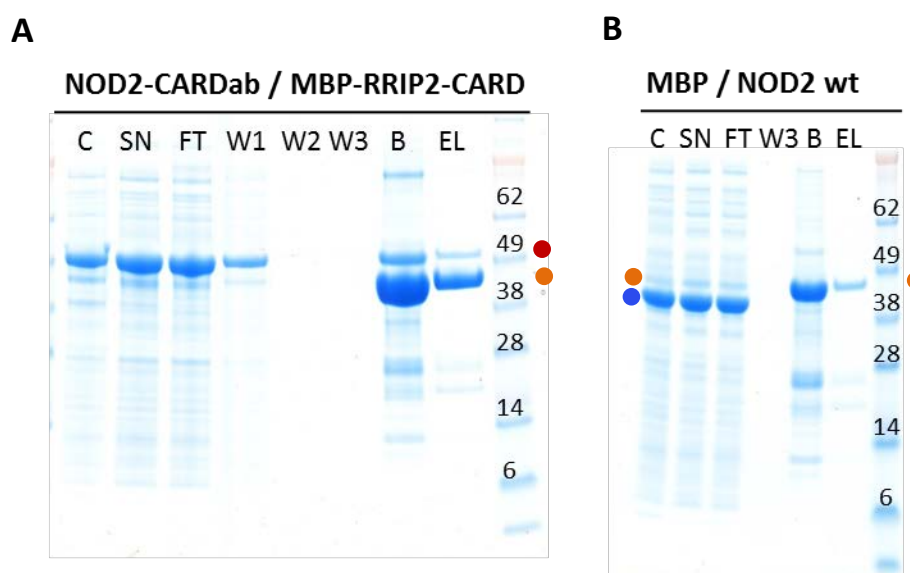


Figure 4-11 Control gels of GST-pull-down experiments. (A) Co-expression and pull-down analysis using glutathione sepharose beads. MBP-RIP2-CARD (52 kDa, red) interacts with GST-NOD2 CARDab (48.5 kDa, orange) on the beads and both proteins can be eluted. (B) The MBP-tag (40 kDa, blue) does not bind GST-NOD2-CARDab. Both proteins are present in the supernatant following co-expression but only GST-NOD2 CARDab is present in the beads and elution sample. C: cell, SN: supernatant, FT: flow through, W1-3: washing steps, B: beads, EL: elution.

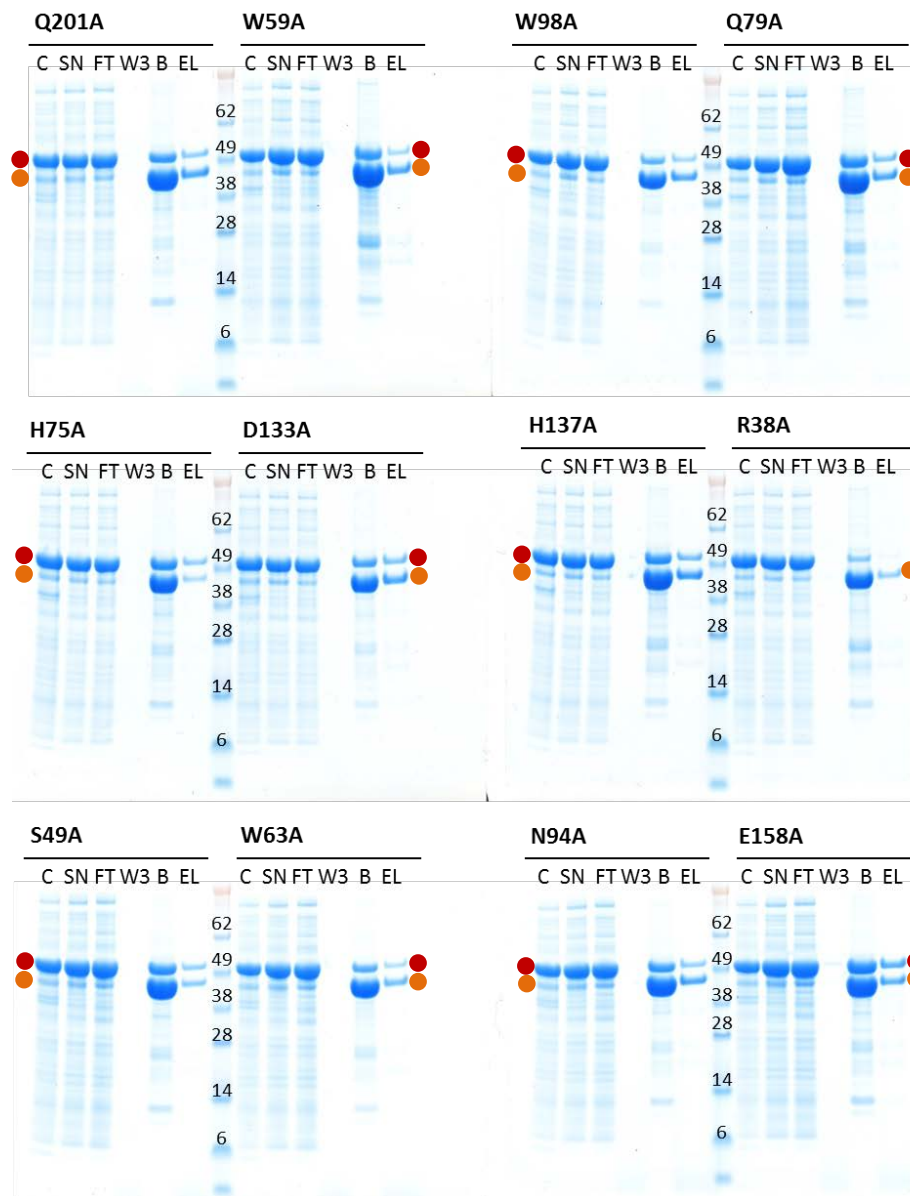


Figure 4-12 GST-pulldowns: Effect of NOD2 CARDab mutants on RIP2 binding. Co-expression of MBP-RIP2-CARD with NOD2-CARDab containing point mutations (CARDa: R38A, S29A, W59A, W63A, H75A, Q79A, N94, W98A. CARDb: D133A, H137A, E158A, Q201A) MBP-RIP2-CARD interacts with GST-NOD2 CARDab on the beads and both proteins can be eluted. Bands corresponding to GST-NOD2 CARDab (48.5 kDa) are marked orange and bands corresponding to MBP-RIP2-CARD (52 kDa) are marked red. C: cell, SN: supernatant, FT: flow through, W3: third washing step, B: beads, EL: elution.

4.4 NOD2-CARDab and RIP2-CARD polymerisation study

It has been recently shown by electron microscopy and x-ray crystallography that CARDS can form a helical filamentous assembly. Polymerisation was quantitatively assessed in that study *via* fluorescence polarisation assays. Additionally, these assays were used to show that an upstream regulator can enhance polymerisation by nucleating these filaments. In these experiments mentioned, polymerisation was initiated by cleaving the CARD from its MBP-fusion protein [103].

The goal of this part of the project was to set up a fluorescence polarisation assay to test if RIP2-CARDS may undergo filament assembly when cleaved off its MBP-tag. Additionally, we wanted to test if the tandem CARDS of NOD2 can initiate and/or enhance RIP2 polymerization.

4.4.1 Constructs, expression and purification

Our MBP-RIP2 aa432-529 construct, which has been shown to be soluble and able to bind NOD2-CARDab, had to be re-cloned using the LIC-vector pHMBP, which was designed in-house. This plasmid generates the desired protein construct including an N-terminal His₆-MBP-tag with a 3C cleavage site for tag-removal. Furthermore, the protein needed to be labelled with a fluorophore-maleimide at a location where the probe would not sterically interfere with any CARD-CARD interface. Hence, labelling on the CARD itself needed to be avoided. For this reason, C455, which is located at the CARD, was changed to an alanine via site-directed mutagenesis. However, to include a cysteine in our construct, two different approaches were taken: firstly, one construct was extended on the N-terminus to aa 413-529 to include the natural occurring cysteine C414 of the RIP2 sequence; secondly, another construct was designed by

extending the fragment to the RIP2 C-terminus at aa 540 instead and inserting the mutation S539C. The newly designed constructs needed to be tested if they were still soluble and if NOD2-CARDab binding can still be achieved. The same pull-down experiment as described above in section 4.4.1.2 were performed with the MBP-RIP2-CARD C455A (aa 413-529) and MBP-RIP2-CARD C455A/S539C (aa 432-540). SDS-PAGE analysis is shown in Figure 4-13. Although both RIP2 constructs are on the beads, only MBP-RIP2-CARD C455A (aa 413-529) is eluted together with NOD2-CARDab. For this reason, this construct was chosen for the following fluorescence polarisation assays. Larger scale purification was performed via Amylose resin and size exclusion chromatography following the standard protocol described in section 2.2.8 and 2.2.10.

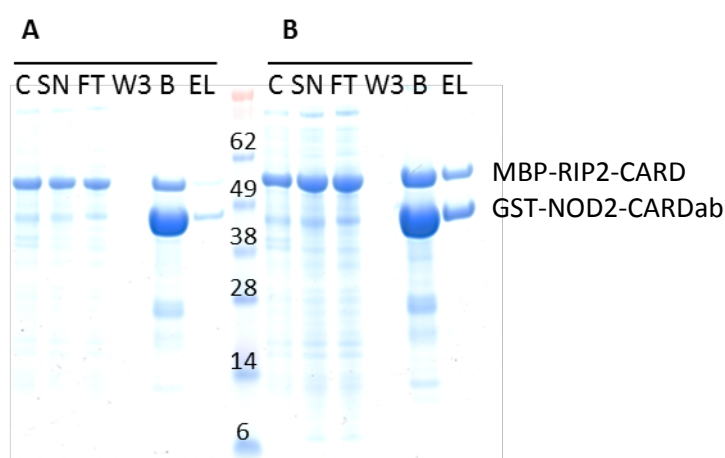


Figure 4-13 GST-pulldowns with modified MBP-RIP2-CARD constructs. (A) MBP-RIP2-CARD C455A/S539C (aa432-540) with GST-NOD2 CARDab. (B) Co-expression of MBP-RIP2-CARD C455A (aa413-529) with GST-NOD2-CARDab and although both RIP2 constructs are on the beads only MBP-RIP2-CARD C455A (aa 413-529) is eluted together with NOD2-CARDab. C: cell, SN: supernatant, FT: flow through, W3: third washing step, B: beads, EL: elution.

4.4.2 Fluorophore labelling

The fluorophore used in this study was Alexa Fluor® 488 C₅ maleimide (Thermo Scientific), a green fluorescent dye designed for conjugation to a thiol group. The protein sample was concentrated to 50 μ M prior to mixing with a 3-fold molar excess of Alexa Fluor® 488 C₅ maleimide. After 2 h incubation at room temperature with the MBP-RIP2 construct, the sample was loaded onto a disposable PD-10 desalting column before another gel filtration chromatography step was performed using 150 mM NaCl, 20 mM Tris-HCl pH 7.5 and 0.5 mM TCEP to remove free dyes (Figure 4-14). The following monomeric peak containing the labelled RIP2 was collected and concentrated to \sim 10 μ M for further studies.

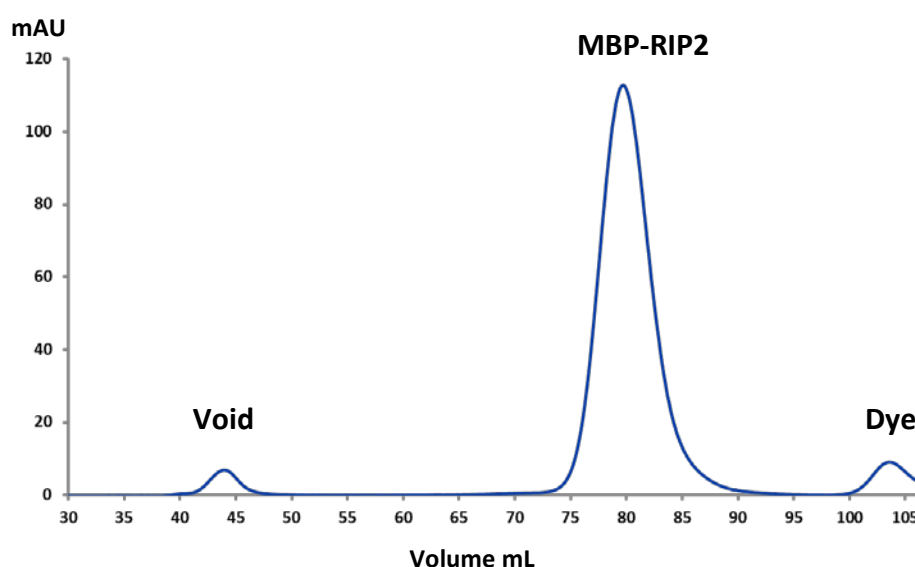


Figure 4-14 His₆-MBP-(3C)-RIP2-CARD C455A (aa 413-529) gel filtration purification after labelling. Elution profiles of the Alexa Fluor® 488 C₅ labelled His₆-MBP-(3C)-RIP2-CARD C455A (aa 413-529) using a Superdex 200 16/60 gel filtration column. The peak of monomeric MBP-RIP2 was collected and pooled. The buffer used was 150 mM NaCl, 20 mM Tris-HCl pH 7.5, 0.5 mM TCEP.

4.4.3 Biophysical analysis

4.4.3.1 UV spectrum

To determine the protein concentration as well as the degree of labelling, a UV spectrum was recorded to measure the absorbance of the conjugate solution at 280 nm and 495 nm. The concentration for the labelled protein was calculated using equation provided by Alexa Fluor® 488 supplier (Thermo Scientific):

Equation 5.1: Protein concentration = $[A_{280} - 0.11(A_{494})] / [\epsilon_{280}(\text{Protein})]$

In this equation, 0.11 is a correction factor for the fluorophore's contribution to the A_{280} .

An absorption of 0.62 at 495 nm and 0.73 at 280 nm was measured. With an extinction coefficient of $72,000 \text{ cm}^{-1}\text{M}^{-1}$ and $74,830 \text{ cm}^{-1}\text{M}^{-1}$ for Alexa Fluor® 488 C₅ and MBP-RIP2 respectively, a protein concentration of 8.9 μM was calculated and a dye concentration of 8.6 μM . Thus it can be concluded that >93 % of the cysteines have been labelled.

4.4.3.2 CD spectroscopy

Before the fluorophore labelling process, it was shown in section 4.4.1 that the MBP-RIP2 (aa413-529) is functional and able to pull-down NOD2-CARDab. To ensure this construct remained folded even after the labelling process, the sample of the labelled MBP-RIP2 was used for CD spectroscopy to determine the secondary structure content. The MBP-RIP2 construct is still clearly alpha-helical (Figure 4-15). The intensity

of this spectrum, however, is probably affected by possible errors in the estimate of protein concentration due to the fluorophore (see 4.4.3.1).

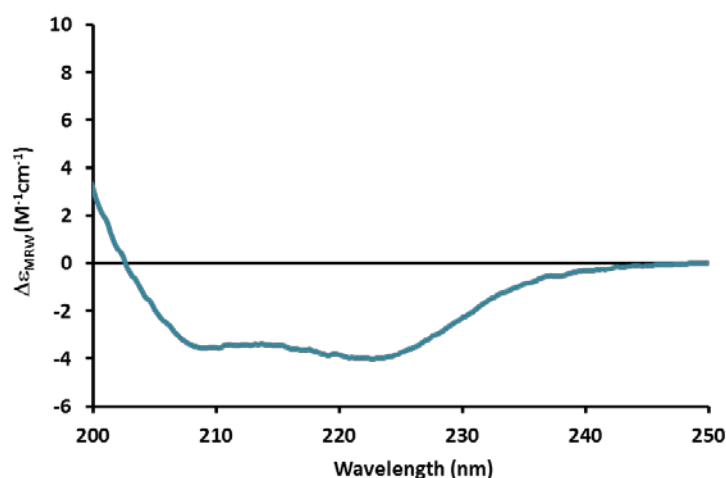


Figure 4-15 Far-UV CD spectra of labelled His₆-MBP-(3C)-RIP2 C455A (aa 413-529). Far-UV CD (220-250 nm) was carried out to analyse the secondary structure content of the labelled His₆-MBP-(3C)-RIP2. Protein concentration was ~0.15 mg/mL and spectra recorded at 20 °C.

4.4.3 Fluorescence polarisation assay with RIP2 and NOD2

Fluorescence polarisation assays as described by Qiao *et al.* [103] was used to investigate potential polymerisation of RIP2 in the presence or absence of NOD2. In preparation for the fluorescence polarisation assay, the excitation and emission spectra of the Alexa Fluor® 488 labelled MBP-RIP2 were recorded (Figure 4-16). Both excitation and emission maxima are close together and we decided on using an excitation of 492 nm and emission of 522 nm for the assay. Fluorescence polarization assays were performed at 25 °C in a buffer containing 150 mM NaCl, 20 mM Tris-HCl pH 7.5 and 0.5 mM TCEP. The sample volume was 80 µL with a MBP-RIP2 concentration of 9 µM. About 100 seconds after starting the experiment with a recording time of 2h (7200 seconds) 5 µg of His₆-3C protease (purified in-house) was

added to the cuvette for removal of the MBP-fusion to allow RIP2 polymerization (Figure 4-17). The 3C cleavage was clearly monitored with a decrease in fluorescence polarisation as the molecular weight of the labelled protein drops from 55.5 to 13.5 kDa resulting in faster tumbling molecules, hence orientational distribution becomes randomized prior to emission, resulting in lower fluorescence polarization [154]. The FPA experiment was repeated, but with the addition of 2 μ M (end concentration) NOD2-CARDab in an NOD2-induced RIP2 polymerization experiment (Figure 4-18 blue). However, a polymerisation of RIP2-CARD following 3C cleavage could not be recorded either in presence or absence of NOD2-CARDab molecules (Figure 4-17 and Figure 4-18 blue). As it has been suggested in the literature that CARD9 may also interact with the RIP2-NOD2 complex [78, 80], an additional FPA experiments was performed, similar to the previous ones but by adding 2 μ M of CARD9 aa 1-277 (Figure 4-18 red). However, the result was similar and no polymerisation of RIP2 could be monitored. After the experiments the samples were analysed via SDS-PAGE. As seen on the gel, all MBP-RIP2 has been cleaved during the experiments (Figure 4-19).

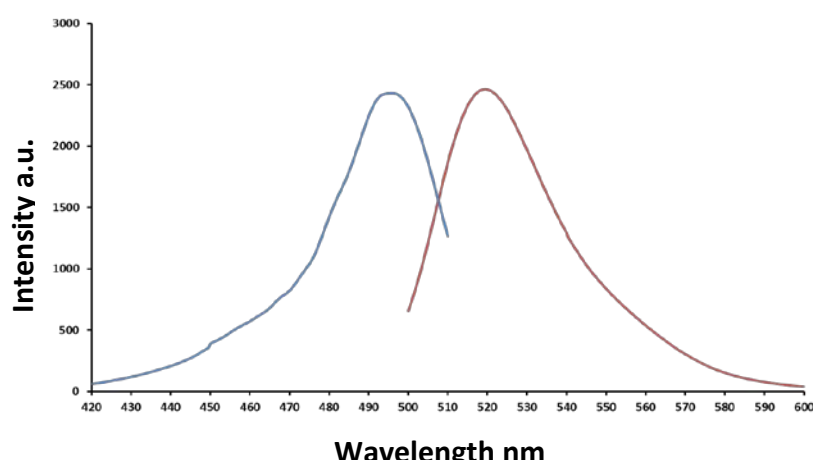


Figure 4-16 Fluorescence spectra. Excitation and emission spectra of Alexa® Fluor 488 labelled His₆-MBP-(3C)-RIP2-CARD C455A (aa 413-529).

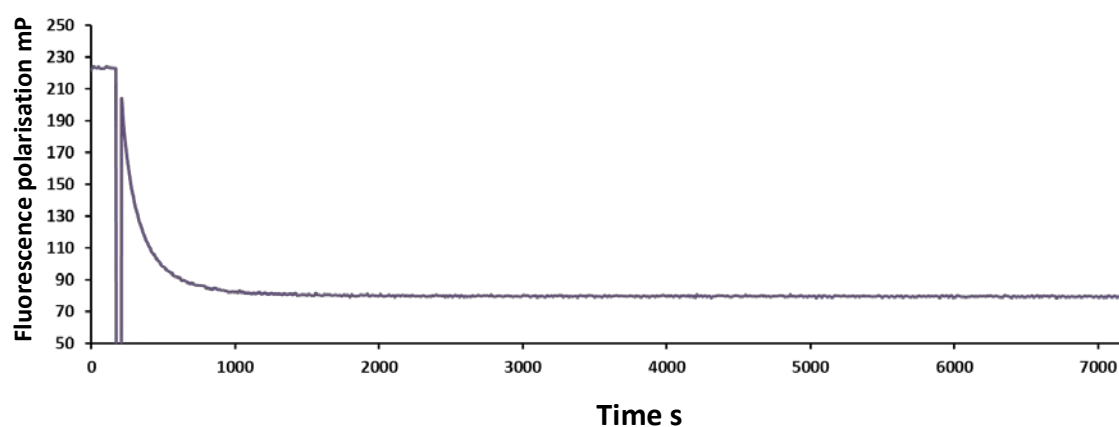


Figure 4-17 Fluorescence polarization assay 1. The fluorescence polarisation of Alexa 488-labeled His₆-MBP-(3C)-RIP2-CARD C455A (aa413-529) was monitored before and after removal of the MBP-tag by 3C protease cleavage. The temporary drop in polarization at the beginning of the experiment represents the addition of 3C protease.

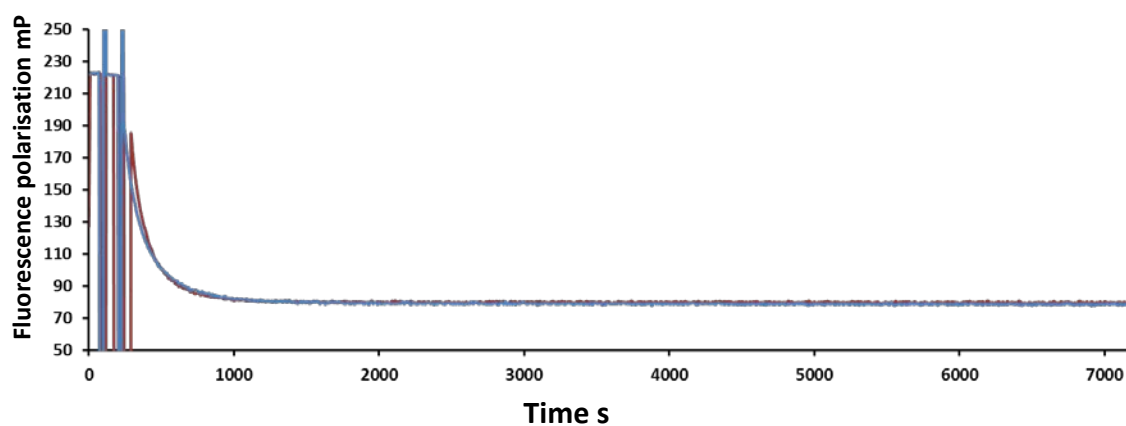


Figure 4-18 Fluorescence polarization assay 2. Blue: Alexa 488-labeled His₆-MBP-(3C)-RIP2-CARD C455A (aa413-529) in the presence of substoichiometric amounts (1/5) of NOD2-CARDab upon MBP removal by 3C protease cleavage. Red: Same experimental setup as (Blue) but in the presence of substoichiometric amounts (1/5) of CARD9 aa 1-277 in addition to NOD2.

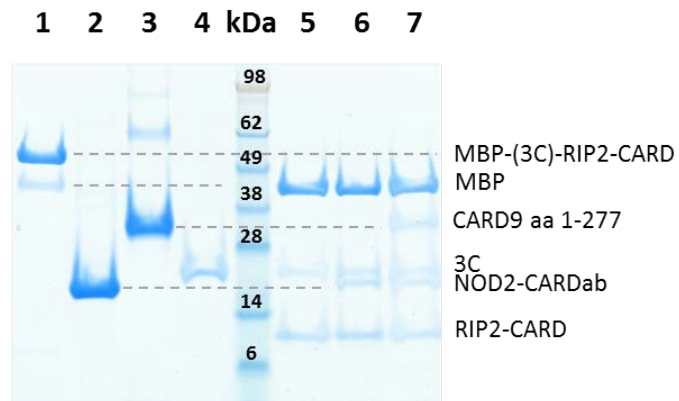


Figure 4-19 SDS-PAGE analysis pre- and post-FPA. 1: uncleaved labelled MBP-RIP2-CARD (56 kDa), 2: NOD2-CARDab (22 kDa), 3: CARD9 aa 1-277 (34 kDa), 4: 3C protease (21.7kDa), 5: post FPA sample - only MBP-RIP2, 6: post FPA sample with added NOD2-CARDab, 7: post-FPA sample with NOD2 and CARD9.

4.5 Summary

In this chapter, preliminary work from former group members has been used as a basis to further define the binding interface of both, the intramolecular interaction of NOD2 tandem CARDs and the intermolecular interaction of NOD2-CARDa with RIP2-CARD.

Expression problems with the NOD2 tandem CARDs, especially mutant proteins, in medium other than LB, forced us to abandon NMR titration experiments. Mutational analysis was therefore continued using ITC. I was able to show that NOD2-CARDa mutants W59A, W63A and S49A were unable to bind CARDb. Comparing the binding surface of NOD2-CARDa involving these three residues, the interaction with CARDb resembles the intramolecular interaction between DED1 and DED2 of MC159 or Caspase-8. However, no residues could be identified in NOD2-CARDb that interfered with the intramolecular interaction.

Previous pull-down experiments revealed that R38 and R86 in NOD2-CARDa are essential for RIP2 binding. All new NOD2-CARDa mutants generated for the ITC analysis were then used for pull-down experiments with MBP-RIP2-CARD in order to further define the binding interface between NOD2 with RIP2. However, all new mutants showed no significant effect and were able to pull-down RIP2-CARD. This data is consistent with previous studies by former group members that suggested that it is mainly NOD2-CARDa that recognises RIP2-CARD.

Due to the discovery that CARDs can form filamentous assemblies that play a signalling role, fluorescence polarisation assays were carried out with RIP2-CARD to investigate its ability to polymerise upon MBP-tag removal. However, no polymerisation could be detected in the absence or the presence of NOD2-CARDa.

Chapter 5

Study of CARMA1, CARD9 and BCL10

5.1 Overview

CARMA1 and CARD9 are both adaptor proteins containing an N-terminal CARD followed by a coiled-coil domain. They are key in initiating the formation of the CBM signalosome for NF- κ B activation and MAPK in lymphoid cells (CARMA1) and myeloid cells (CARD9). CBM-complex assembly involves the recruitment of BCL10 via a CARD-CARD interaction to further complete the signalosome by binding to MALT1.

At the time when we discovered that the construct of CARMA1-CARD is highly soluble (as seen in Chapter 3), no structures of CARMA1, CARD9 or BCL10 were available. Therefore, the original aim of the work described in this chapter was to structurally characterise these individual CARD domains and to biophysically investigate the CARD-CARD binding modes involved in forming a complex between CARMA1 and BCL10 as well as CARD9 and BCL10. Those two systems, strikingly similar from a structural point of view, yet so different in a cellular context, make the comparison between the characteristics of CARMA1 and CARD9 even more intriguing. Additionally, a characterisation of CARMA1 and CARD9 and how they interact with BCL10 could pave the way towards the final goal of finding a general mode of interaction between CARD domains.

5.2 CARMA1

This section explores the two approaches used to characterise CARMA1. Firstly the attention was focused on the production, purification and characterisation of the CARMA1-CARD domain in isolation, with the goal to determine its biophysical properties and ultimately its structure. In the second stage of the study, the focus was turned to constructs comprising the N-terminal CARD followed by different segments of the coiled-coil domain. This approach is based on the rationale, that CARD-containing proteins may require an additional domain, which undergoes oligomerisation to enable the interaction with its binding partners. Therefore, the coiled-coil segments may provide a platform for BCL10 recruitment.

5.2.1 CARD of CARMA1

5.2.1.1 Expression and purification

The results from the small scale solubility screen (see section 3.2.2) originally drew our attention to CARMA1-CARD. These experiments showed that the CARD of CARMA1 containing aa 18-110 is well expressed either as GST-fusion or as a His₆-tagged protein. It was also possible to purify CARMA1-CARD aa 18-110 from a larger expression using 750 mL LB media, performing either a Ni-NTA purification for the His-tagged protein or GST-affinity chromatography for the GST-fusion construct. In both cases, soluble CARMA1-CARD was obtained after elution using 500 mM Imidazole (for the His-tagged protein) or 3C cleavage (for the GST-tagged protein) (see Figure 3-7 and Figure 5-1 respectively). For this reason, CARMA1-CARD was chosen for further biophysical investigation as it was one of the most soluble constructs of the expression screens and – in contrast to other CARDS like APAF1, RAIDD or NOD1, which also showed high

solubility – there was nothing known about the structure of CARMA1 at that time. Both CARMA1-CARD samples (after either GST-affinity chromatography or Ni-NTA purification) were further purified via gel filtration chromatography to remove any remaining contaminants. The pure protein samples could be concentrated up to 20 mg/ml without showing any signs of precipitation.

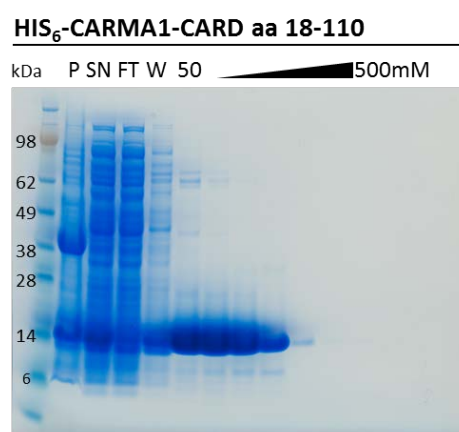


Figure 5-1 His₆-CARMA1-CARD purification. SDS-PAGE analysis following purification of His₆-tagged CARMA1-CARD aa18-110 (13.2 kDa). M; protein marker, P; pellet, SN; supernatant, FT; flow through, W; washing step. The protein was eluted by an imidazole gradient.

5.2.1.2 Crystallisation experiments

Although interested in the structural details of homotypic CARD-CARD interactions, obtaining the crystal structure of CARMA1-CARD would open up opportunities to investigate the molecular details of possible binding sites. Therefore, several initial crystallisation screens were kindly set up by Lesley Haire and Roksana Ogrodowicz; CARMA1-CARD (aa 18-110) obtained after 3C cleavage of the GST-tag was used to set up an initial screens at 17 mg/ml, using the sitting drop method and an incubation temperature of 18 °C (using the following screens: Imperial College screen (designed

by Lesley Haire), Classics lite and pH clear 1 (both Qiagen)). In addition, further screens (pH clear 1 and 2, Classics lite, Wizard 1 and 2 (Rigaku reagents)) were set up using His₆-CARMA1 at 10 mg/mL and 20 mg/mL. I observed that the ratio between precipitation and clear drops was balanced. Unfortunately, no crystals formed after an incubation time longer than 13 days.

5.2.1.3 NMR spectroscopy

In parallel to the crystallisation trials, 1D-¹H NMR spectroscopy was used to analyse the structural integrity of CARMA1-CARD, while 2D and 3D NMR experiments were carried out to assess if structural studies of CARMA1-CARD would be possible by NMR. In these experiments, only the 3C cleaved CARMA1-CARD (aa18-110) construct originally expressed as GST-fusion was used. The protein sample was dialysed against 20 mM HEPES pH 7, 50 mM NaCl, 2 mM TCEP and concentrated to 200 μM. A 1D-¹H NMR spectrum was recorded at 25 °C on a 600 MHz spectrometer (Figure 5-2).

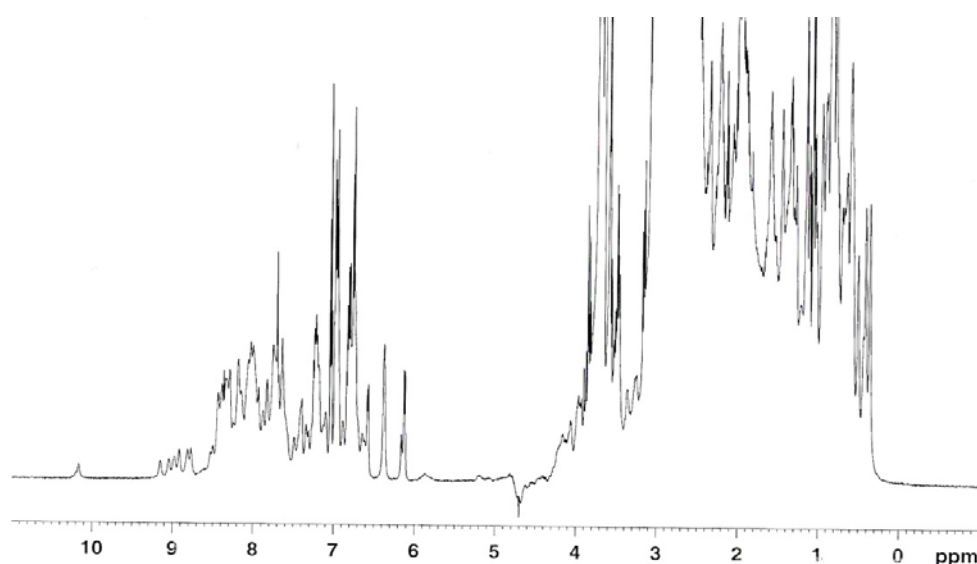


Figure 5-2 1D-¹H NMR spectrum of CARMA1-CARD. NMR analysis shows that CARMA1-CARD contains tertiary structure as peaks at a chemical shifts >9 ppm and ≤0.5 ppm can be observed. The spectrum was recorded at 25 °C on a 600 MHz spectrometer.

Peaks at a chemical shift >9 ppm and ≤ 0.5 ppm indicate that CARMA1 contains a hydrophobic core and is properly folded. To further analyse if structure determination by NMR would be possible, a ^{13}C - ^{15}N labelled CARMA1-CARD sample was produced to carry out heteronuclear multidimensional NMR experiments. The ^{13}C - ^{15}N labelled protein sample was prepared by growing bacteria in 4 litres minimal medium using ^{15}N -ammonium sulphate as well as ^{13}C -glucose. Purification was carried out as for the unlabelled sample and the purified protein was dialysed against 50 mM NaCl, 20 mM HEPES pH 7.0 and 2 mM TCEP before concentrating it to 500 μM . A ^1H - ^{15}N HSQC was recorded at 25 $^{\circ}\text{C}$ which is shown in Figure 5-3.

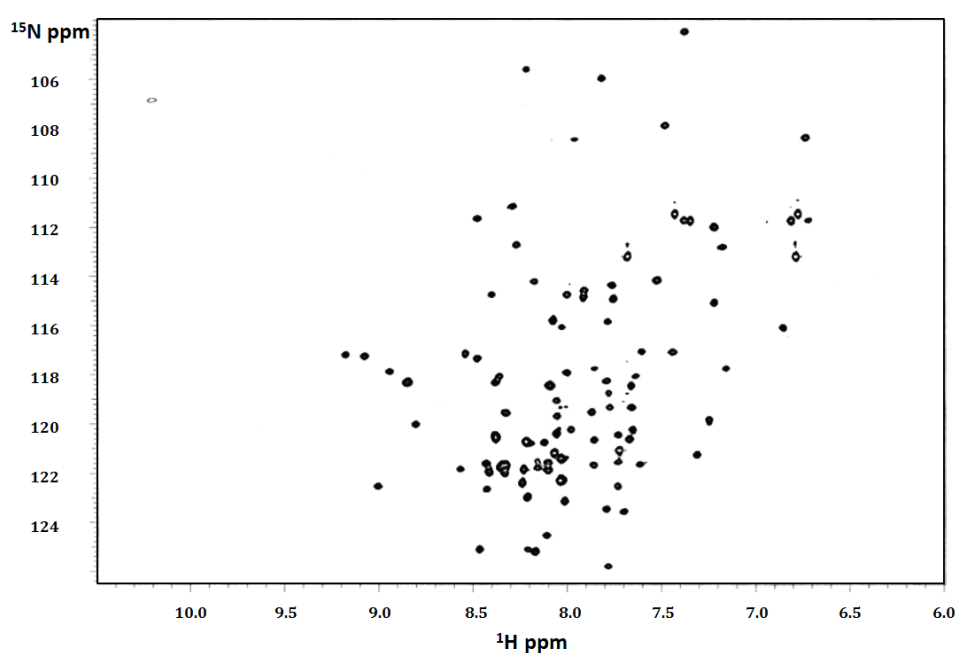


Figure 5-3 ^1H - ^{15}N HSQC spectrum of CARMA1-CARD. The 2D spectrum of CARMA1-CARD shows sharp linewidths of the peaks and confirms the previous result of the 1D proton spectrum, that CARMA1 is properly folded. A protein concentration of 500 μM was used in 50 mM NaCl, 20 mM HEPES pH 7, 2 mM DTT and the spectrum was recorded at 25 $^{\circ}\text{C}$ on a 600 MHz spectrometer.

The spectrum is of very good quality because the peaks are well dispersed; this confirms the previous result obtained from the 1D proton spectrum that CARMA1 is properly folded. With a peak number of 102 (plus 7 prolines in the sequence, which can't be observed in this spectra) nearly all 106 residues are observable. The sharp linewidth of the peaks is consistent with a single species in solution, most likely a monomer. For structure determination, the quality of triple-resonance spectra is essential and for this reason, a short HNCO was recorded (Figure 5-4); The HNCO experiment is the most sensitive multi-dimensional experiment and considerably quick to record, and for this reason is normally chosen as the first spectrum to test a protein sample before running other 3D experiments [155].

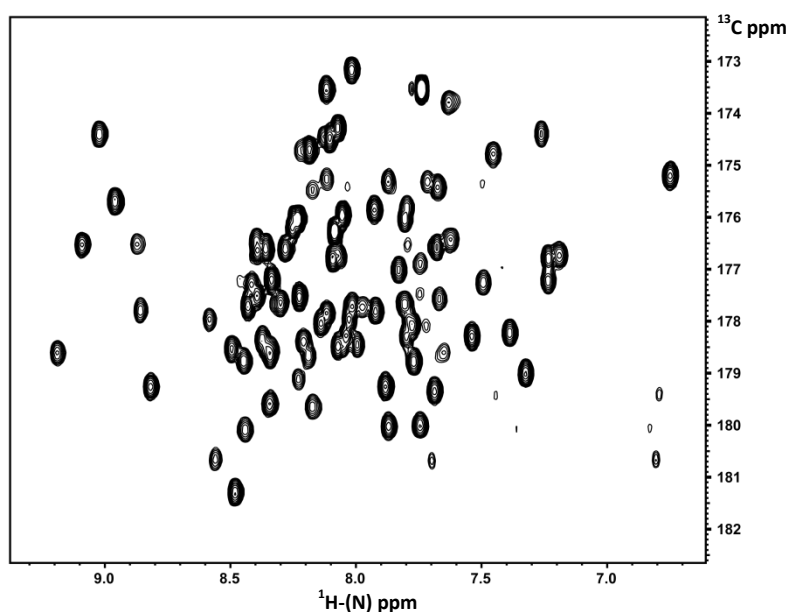


Figure 5-4 ^1H - ^{13}C - ^{15}N HNCO experiment of CARMA1-CARD. The H-C(O) plane of the HNCO shows sharp peak linewidth and a peak count of 90. A protein concentration of 500 μM was used in 50 mM NaCl, 20 mM HEPES pH 7, 2 mM DTT and the spectrum was recorded at 25 $^\circ\text{C}$ on a 600 MHz spectrometer.

Similarly to the HSQC spectrum, the HNCO spectrum also shows sharp peak linewidths and a good peak distribution. An estimated peak count of 90 coincides with the 96 residues of the protein. From those preliminary experiments it can be concluded that an assignment and a structural determination would be possible with a complete set of 3D experiments. Unfortunately, before more crystallisation screens were set up and additional NMR experiments recorded, Li *et al.* (2012) [156] published the crystal structure of CARMA1-CARD and the focus was shifted towards longer CARMA1 constructs and CARD9.

5.2.2 CARD-coiled-coil constructs of CARMA1

5.2.2.1 Construct design

As previously explained, the reasons for producing extended constructs of CARMA1 is based on the rationale that CARD-containing proteins may require an additional domain, which undergoes oligomerisation, to enable the interaction with partner CARDS [54]. As seen in Figure 5-6 A, CARMA1-CARD is predicted to be followed by a coiled-coil domain, which is connected with the CARD by a linker of unknown structure. Importantly, the coiled-coil is an oligomerisation domain, which could be an important feature to enable CARD-CARD interactions. The MultiCoil software [157] was used to predict potential coiled-coil regions in CARMA1 (Figure 5-5). There are four predicted coiled-coil segments starting and ending around residues 130-175, 200-300, 300-350, and 350-430, respectively (Figure 5-6 A). Based on the prediction, 15 constructs have been designed starting at amino acid 18 and ending at various lengths along the coiled-coil domain (Table 5-1).

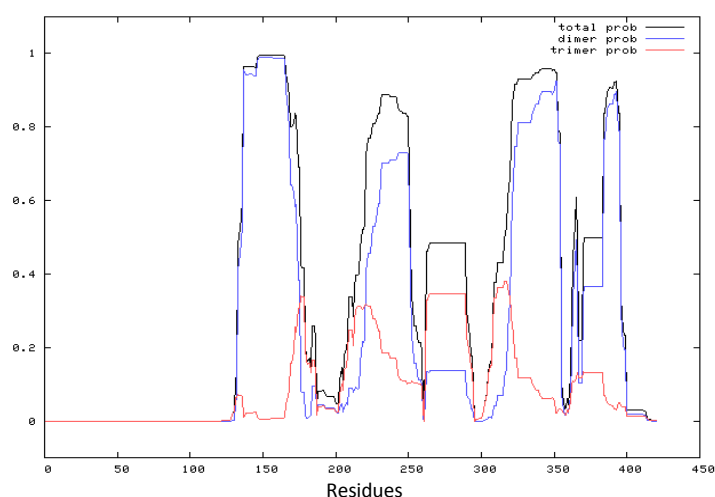


Figure 5-5 Prediction of coiled coil structures by MultiCoil. Approximately four potential coiled coil regions in CARMA1 are shown, around 130-175, 200-300, 300-350, and 350-430 [157].

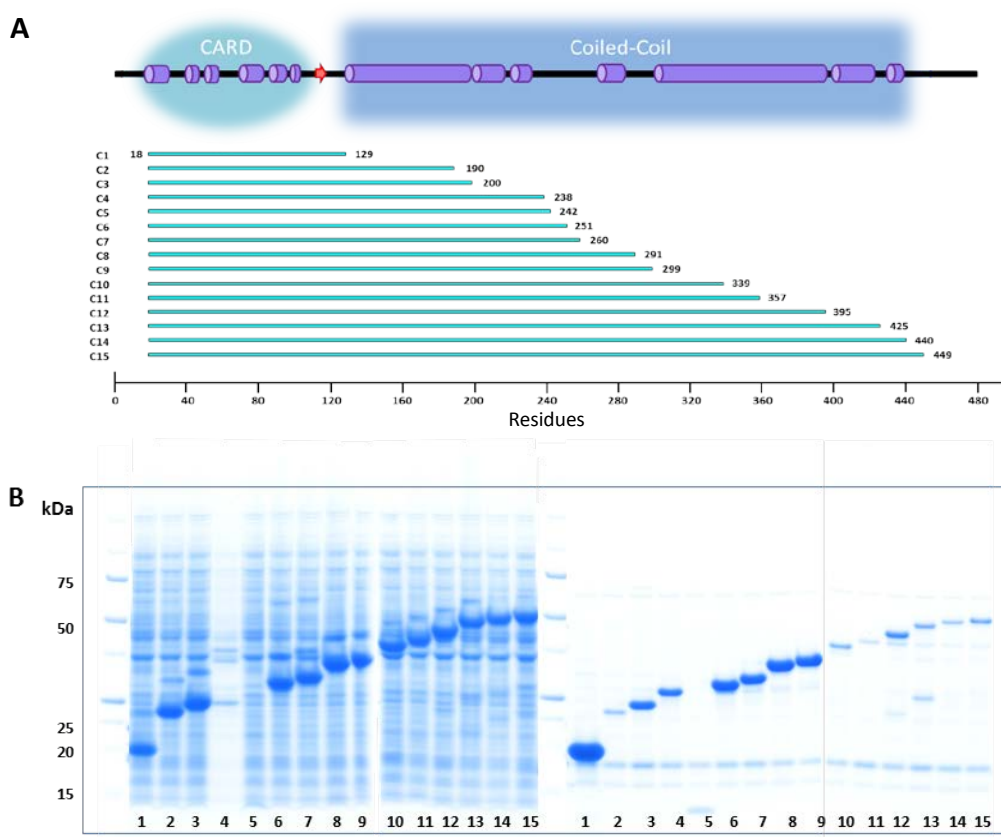


Figure 5-6 CARMA1-CARD-CC constructs. A; Secondary structure prediction (JPred [145]) of the CARMA1 CARD and coiled-coil (CC) domain is illustrated with the designed constructs C1-C15. B; SDS-PAGE gel results of the small scale expression and purification screens of the 15 CARMA1 constructs, with the induced cell lysate on the left, followed by the proteins eluted from the Ni-NTA-column on the right. The constructs have been numbered and listed in Table 5-1.

#	construct	MW [kDa]	pI	expressed	soluble
1	18-129	15.3	5.90	✓✓✓	✓✓✓
2	18-190	22.9	6.21	✓✓✓	✓
3	18-200	24.1	5.95	✓✓✓	✓✓
4	18-238	28.7	6.28	✓	✓✓
5	18-242	29.2	5.92	×	×
6	18-251	30.3	6.03	✓✓✓	✓✓✓
7	18-260	31.4	6.16	✓✓✓	✓✓✓
8	18-291	35.0	6.17	✓✓✓	✓✓✓
9	18-299	35.9	6.17	✓✓✓	✓✓✓
10	18-339	40.8	5.76	✓✓✓	×
11	18-357	42.8	5.56	✓✓✓	×
12	18-395	47.5	5.49	✓✓✓	✓
13	18-425	51.4	5.53	✓✓✓	✓
14	18-440	53.1	5.71	✓✓✓	✓
15	18-449	54.1	5.59	✓✓✓	✓

Table 5-1 CARMA1-CARD-CC constructs. 15 constructs of CARMA1, including the CARD domain and additional amino acids on the C-terminal domain, have been screened by small scale expression and purification experiments. Results are illustrated as heavily ✓✓✓, moderately ✓✓ and minimally ✓ expressed/soluble constructs based on the visual analysis of SDS-PAGE gels. Unexpressed/insoluble constructs are marked with ×.

All constructs have been cloned into pET52-47 to obtain N-terminally His₆-tagged proteins. To screen the newly generated CARMA1 constructs efficiently, a small scale expression and solubility screen with Ni-NTA was carried out as described in section 2.2.4. Most of the constructs containing fractions of the first two predicted coiled-coil segments (up to residue 299) were soluble.

To be able to determine the function of the additional coiled-coil segments, purification of the constructs aa 18-395 and aa 18-449, both containing the full coiled-coil, was attempted by modifying the lysis buffer (300 mM NaCl, 50 mM L-Arginine/Glutamic acid, 50 mM Tris-HCl pH 8, 10 mM Imidazole, 0.3 mM TCEP, 10 % glycerol) to possibly enhance protein stability/solubility. However, SDS-PAGE gel

analysis of the different purification steps showed that most of the protein still remained insoluble or irreversibly bound to the Ni-NTA beads (Figure 5-7). The small amount of protein finally eluted with 500 mM Imidazole was not suitable for further purification steps due to protein precipitation.

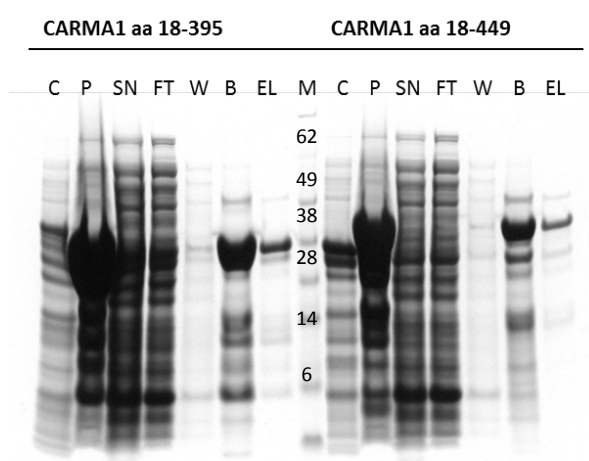


Figure 5-7 His₆-CARMA1 purification. SDS-PAGE analysis following expression of His₆-tagged CARMA1 aa18-395 (47.5 kDa) and aa 18-449 (54.1kDa). M; protein marker, P; pellet, SN; supernatant, FT; flow through, W; washing step, B; Beads and EL; elution.

In addition, considering the possibility that a different tag could enhance solubility, constructs CARMA1 aa 18-440 and aa 18-449 were cloned into pET52 to express STREP-tagged proteins. Purification with the above mentioned modified lysis buffer (excluding Imidazole) was attempted using StrepTactin sepharose beads (GE Healthcare) and 2.5 mM desthiobiotin for elution. As this approach did not lead to any soluble proteins either, a 24-well small scale solubility screen using different lysis buffers with a variation of salts, pH and additives such as glycerol, trehalose or triton X-100, was performed and the lysate analysed via SDS-PAGE. However, even in this

case no improved condition was found (data not shown). For this reason, the focus was then set on the truncated constructs containing only up to 2 of the 4 coiled-coil segments to further analyse them via SEC-MALLS.

5.2.2.2 SEC multi-angle laser light scattering

According to the MultiCoil prediction in Figure 5-5, the CARMA1-coiled-coil most likely forms a dimer. To investigate the influence of the different coiled-coil fractions on protein oligomerisation, we used multi angle laser light scattering coupled to size exclusion chromatography (SEC-MALLS) to investigate the molecular masses of selected constructs in solution. The constructs chosen for these experiments were CARMA1 aa 18-129, 18-200 and 18-299. They have been successfully purified following a standard Ni-NTA purification protocol (2.2.6) including 10% glycerol. The purified proteins were concentrated to 3 mg/mL before loading onto a S200 10/300 GL column equilibrated in 150 mM NaCl, 50 mM HEPES pH8, 1mM EDTA, 1 mM TCEP and 3 mM Azide at 4 °C. For the construct CARMA1 aa 18-129, SEC-MALLS analysis resulted in a main peak with a recorded molecular weight of 14.7 kDa consistent with a monomer (calculated M_w = 15.3 kDa). A small amount (< 7%) of the protein forms a dimer, showing up as a second peak with a molecular weight of 30.1 kDa (Figure 5-8). For CARMA1 aa 18-200, only one peak was monitored resulting in a molecular weight of 46.5 kDa (calculated M_w for the monomer = 24.1 kDa) which is clearly forming a dimer. The third sample, CARMA1 aa 18-299, unfortunately eluted mainly in the void and could not be further analysed. In this case, purifying and analysing the next shorter CARMA1 construct (aa 18-291) could provide more insights.

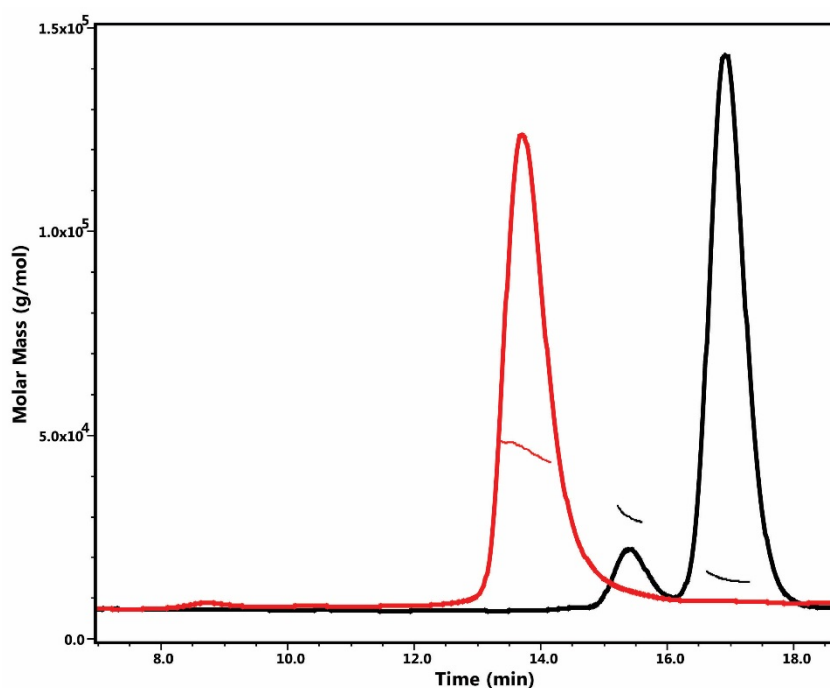


Figure 5-8 SEC-MALLS analysis of CARMA1. The chromatogram shows the UV absorbance and the molecular weight calculation of the samples. The samples were loaded at 3 mg/mL. Black; SEC-MALLS analysis of CARMA1 aa 18-129, resulted in a main peak with a recorded molecular weight of 14.7 kDa consistent with a monomer (calculated $M_w = 15.3$ kDa). Second peak corresponds to a dimer with a molecular weight of 30.1 kDa. Red; For CARMA1 aa 18-200, only one peak was monitored resulting in a molecular weight of 46.5 kDa (calculated M_w for the monomer = 24.1 kDa).

5.3 CARD9

5.3.1 CARD of CARD9

5.3.1.1 Constructs, expression and purification

CARD9 also comprises an N-terminal CARD followed by a coiled-coil similar to CARMA1. More interestingly, it has been shown that CARD9 also recruits BCL10 via CARD-CARD interaction in myeloid cells. Therefore, the His₆-tagged CARD9 aa 1-104, which appears to be at least partly soluble in the small scale expression and solubility screen (see section 3.2.2) was expressed in 750 mL LB medium and purified using Ni-NTA beads to investigate if we could obtain a reasonable quantity of CARD9 for further experiments (Figure 5-9). Unfortunately, it became soon clear that most of CARD9 is insoluble. Protein from the supernatant, which was loaded on the Ni-beads and eluted was collected but quickly precipitated during the attempt to concentrate it for a further purification step.

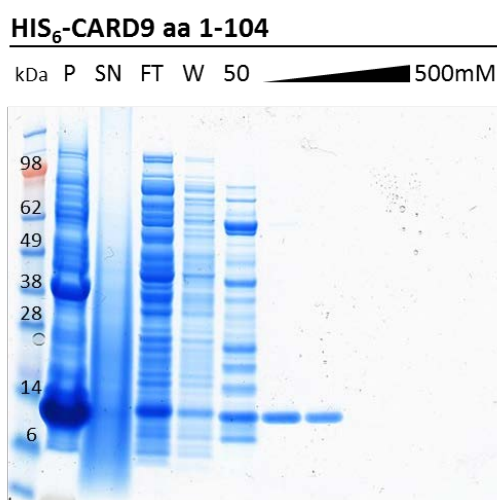


Figure 5-9 His₆-CARD9-CARD purification. SDS-PAGE analysis following expression of His₆-tagged CARD9-CARD aa1-104 (12 kDa). M; protein marker, P; pellet, SN; supernatant, FT; flow through, W; washing step. The protein was eluted with an imidazole concentration of 75 mM and increasing to 500 mM.

Having found that CARD9 aa 1-104 is not a suitable construct to work with, further cloning was carried out. Taking into account the fact that the boundaries for the previous construct (aa 1-104) were set generously to safely include the full CARD, there was still room to both shorten or lengthen the new constructs at the C-terminus. Therefore, three more (aa 1-98, aa 1-122 and aa 1-129) constructs were designed and cloned into pET52-47 to express His-tagged proteins. Upon successful cloning, the proteins were immediately expressed in a larger scale of 750 mL LB medium. The purification was performed as described in 2.2.6. Figure 5-10 shows that all constructs are soluble. For CARD9 aa 1-98 a further purification step was carried out via size exclusion chromatography. The fractions collected were pooled and concentrated. Unfortunately the protein precipitated during the process.

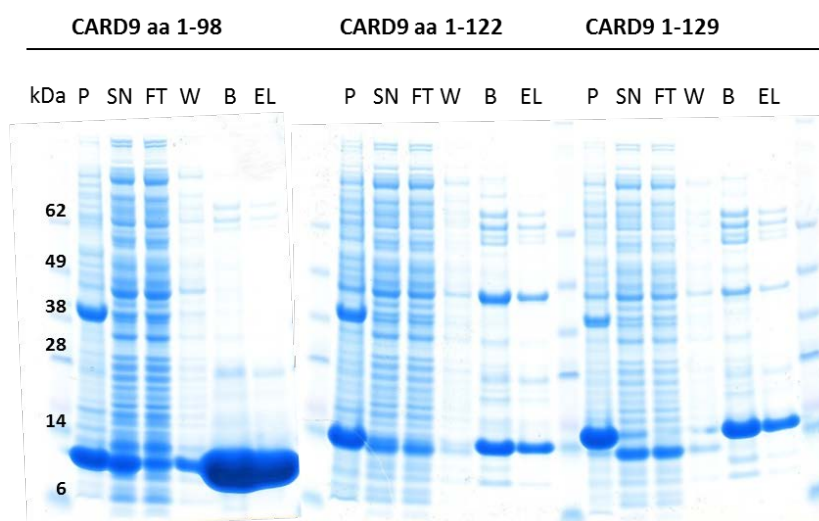


Figure 5-10 His₆-CARD9-CARD purification. SDS-PAGE analysis following expression of His-tagged CARD9-CARD aa 1-98 (13.6 kDa), aa 1-122 (16.1 kDa) and aa 1-129 (17 kDa). P; pellet, SN; supernatant, FT; flow through, W; washing step, B; beads, EL; elution. The protein was eluted with an imidazole concentration of 500 mM.

5.3.1.2 CARD9-CARD aa 1-98: Thermofluor

At this point, it had to be taken into account that we are not only looking for soluble constructs, but also for proteins, which were suitable for crystallisation trials. This means, smaller constructs are more likely to lack flexible and unstructured N- and C-termini, which could affect the propensity to crystallize [158]. For this reason, the focus was directed towards CARD9 aa 1-98, the shortest construct available so far. To identify buffer conditions that would allow CARD9 aa 1-98 to be highly concentrated (at least > 5mg/mL) for crystallisation experiments, the purified protein sample was used for Thermofluor analysis. This fluorescence-based experiment is a screen to search for buffer conditions or additives under which the protein shows maximum thermal stability. Stability was measured by thermal unfolding of the protein and determination of the different melting points (T_m), which can vary considerably depending on the specific conditions. The protein was mixed with the fluorophore SyPro Orange (Thermo Scientific) and added to two 96-well screens. The screens tested were the commercially available Solubility & Stability and Slice pH both from Hampton Research. The data were analysed with the Prism software where the sigmoidal curve was fitted using the Boltzmann equation in order to determine T_m . Selected results following data analysis are shown in Figure 5-11 and 5-12 as well as Table 5-2 and 5-3. Whereas the top 10 conditions from the Slice pH screen vary in most cases between pH 6.4 and pH 8.5, it becomes clear that CARD9 seems to be most stable in BIS-TRIS propane. Additionally, additives such as sodium bromide, glycerol and lithium chloride also increase stability of CARD9.

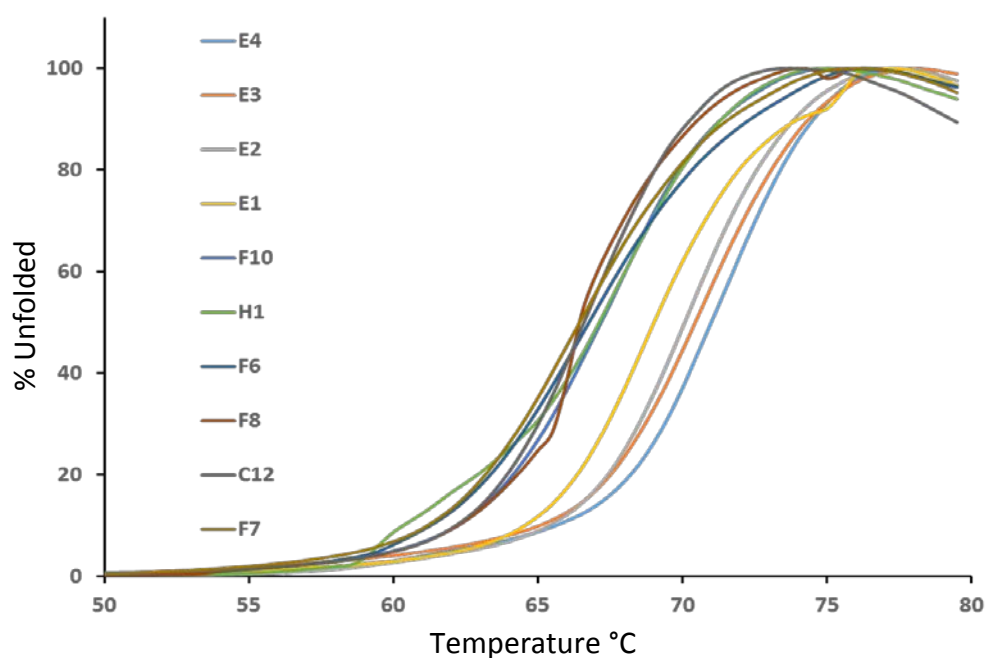


Figure 5-11 Slice pH thermofluor assay with CARD9 aa 1-98. The 10 best conditions are shown for the Slice pH screen. Values of T_m were determined by fitting the sigmoidal curve using the Boltzmann equation and they are listed in Table 5-2 together with the corresponding conditions.

Well	Buffer	T_m
E4	1M BIS-TRIS propane pH 7.3	71.0
E3	1M BIS-TRIS propane pH 7.0	70.5
E2	1M BIS-TRIS propane pH 6.7	70.1
E1	1M BIS-TRIS propane pH 6.4	69.2
F10	1M Tris pH 7.3	67.1
H1	1M BIS-TRIS propane pH 8.5	67.0
F6	1M TRIS hydrochloride pH 7.2	66.8
F8	1M TRIS hydrochloride pH 7.8	66.6
C12	1M BIS-TRIS pH 6.3	66.6
F7	1M TRIS hydrochloride pH 7.5	66.5

Table 5-2 Evaluation of Slice pH thermofluor assays with CARD9 aa 1-98. The 10 best conditions are shown for the Slice pH screen with their melting Temperature obtained by fitting the sigmoidal curve using the Boltzmann equation. This screen shows that the pH can vary between 6.3 and pH 8.5 buffer depending, but it becomes clear that CARD9 aa 1-98 seems to be most stable in BIS-TRIS propane at pH 7.3.

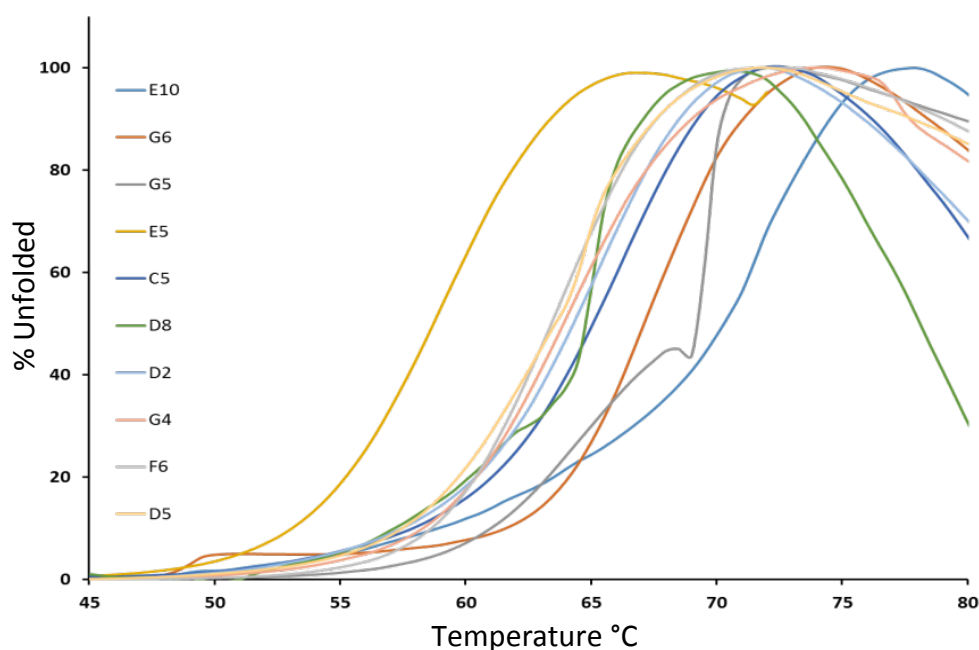


Figure 5-12 Solubility and Stability thermofluor assay with CARD9 aa 1-98. The 10 best conditions are shown for the Solubility and Stability screen. Values of T_m are listed in Table 5-3 together with the corresponding conditions.

Well	Additive	T_m
E10	Ethylammonium nitrate	69.56
G6	Glycerole, Lithium chloride	67.21
G5	Sodium bromide	63.56
E6	Tetraethylammonium bromide	65.22
C5	Adipic acid	64.91
D8	Non Detergent Sulfoetaine 211	64.38
D2	L-Glutathione reduced/oxidized	64.1
G4	Lithium chloride	63.9
F6	Lithium nitrate	63.37
D5	MgCl ₂ , CaCl ₂	63.22

Table 5-3 Evaluation of the Solubility and Stability thermofluor assays with CARD9 aa 1-98.

The 10 best additives are shown for the Solubility and Stability screen together with the obtained melting temperature. Additives such as sodium bromide, glycerol and lithium chloride increase thermal stability of CARD9 aa 1-98.

5.3.1.3 Crystallisation experiments

Based on the results from the thermofluor assays, a new purification protocol was established with a lysis buffer made of 150 mM BIS-TRIS propane pH 7.5, 400 mM LiCl, 50 mM L-Arginine/Glutamic acid, 10 mM Imidazole, 10 % glycerol and 0.2 mM TCEP. The size exclusion buffer was altered to 100 mM BIS-TRIS propane pH 7.5, 250 mM LiCl, 2 mM TCEP and 5 % glycerol. Similar to the purification performed beforehand, CARD9 aa 1-98 was soluble under these condition, and the final concentration could be increased up to 10 mg/ml. Therefore, crystallisation screens were kindly set up by Lesley Haire and Roksana Ogrodowicz for CARD9 (aa 1-98) at 10 mg/mL using the sitting drop method (Index screen (Hampton Research), IC screen (Imperial College screen, designed by Lesley Haire) and Wizard 1 and 2). All plates were incubated at 18 °C but unfortunately no crystals were obtained.

Only initial experiments were set up at that point, because the specific lysis buffer determined via Thermofluor opened up possibilities to focus on longer CARD-9 constructs as described in the next section.

5.3.2 CARD-coiled-coil constructs of CARD9

5.3.2.1 Constructs design

In a similar fashion to what was done for CARMA1, there was the interest to investigate the role of the coiled-coil of CARD9 oligomerisation. In preparation for this, 15 CARD9 constructs were designed ending at various positions along the coiled-coil domain (Table 5-4). In case of CARD9, two coiled-coil prediction software programmes were used to design the constructs: COILS and MultiCoil [157, 159]. There are four predicted coiled-coil segments ending at around residues 180, 230, 380, and 420,

respectively (Figure 5-13). All constructs were cloned into pET52-47 to obtain N-terminally His₆-tagged proteins. To screen the newly generated CARD9 constructs efficiently, a small scale expression and solubility screen with Ni-NTA was carried out as described in 2.2.4. (Figure 5-14) The screen shows that constructs up to residue 277 are soluble including the first two segments of the coiled-coil. Interestingly, this is a result that closely resembles what was found for CARMA1 (Figure 5-6).

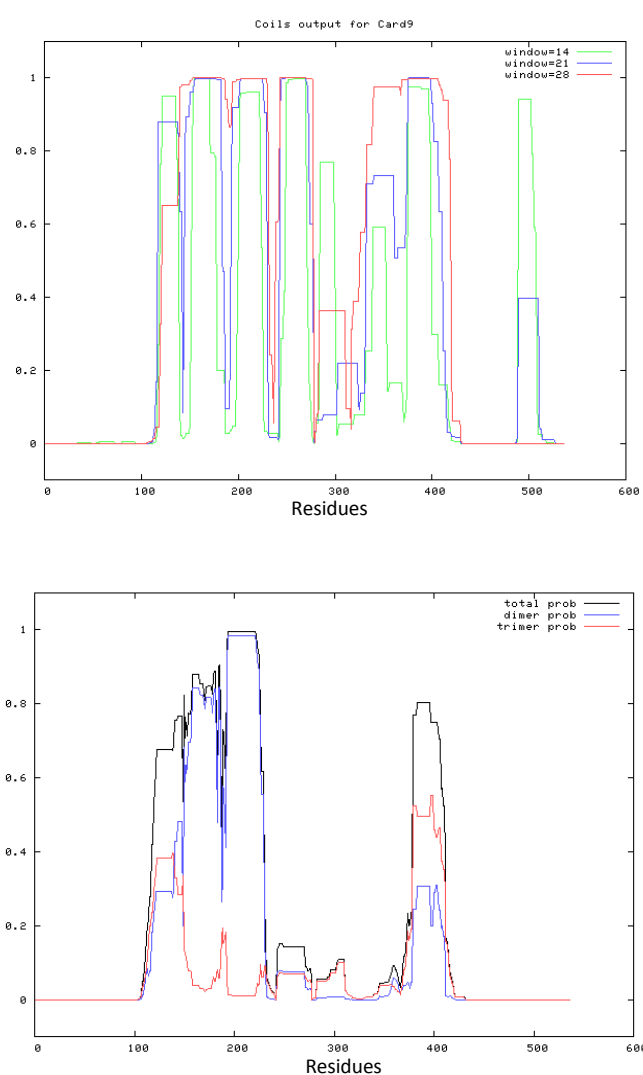


Figure 5-13 Prediction of coiled coil structures by COILS and MultiCoil. Approximate four potential coiled coil regions in CARD9 are predicted using COILS (top) and as a comparison MultiCoil (bottom). Segments are located around 110-180, 180-230, 230-380, and 380-420 [157, 159].

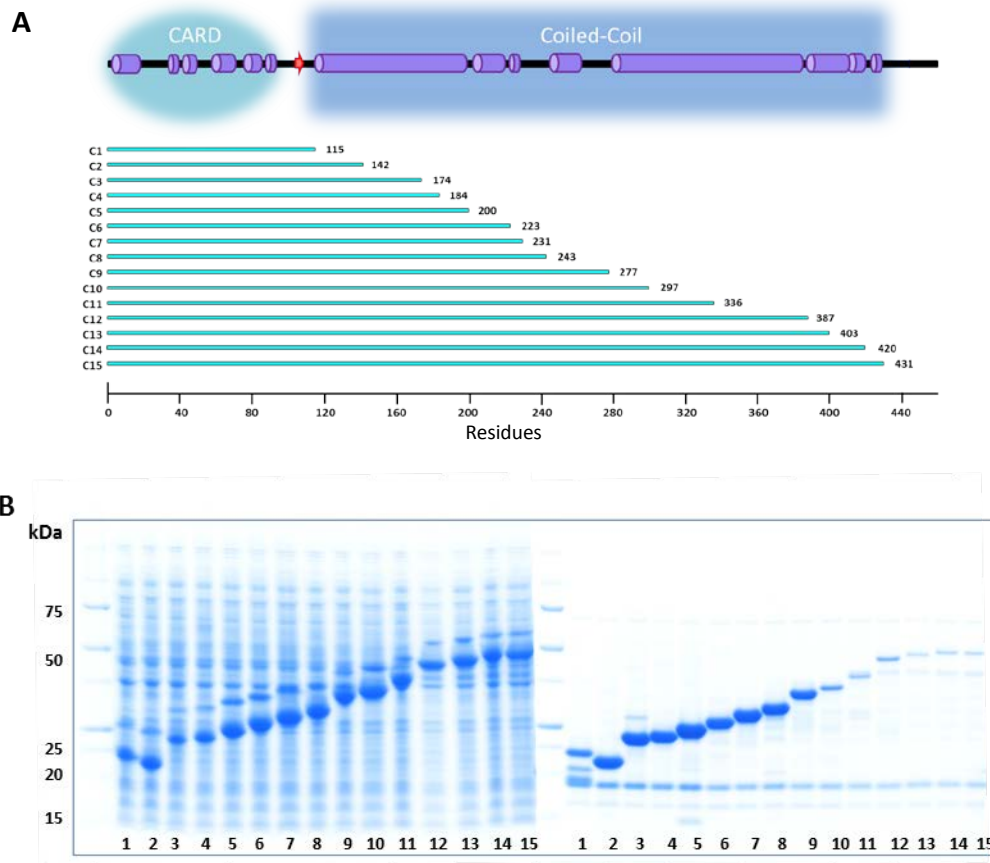


Figure 5-14 CARD9-CARD-CC constructs. A: Secondary structure prediction (JPred [145]) of the Card9 CARD and coiled-coil (CC) domain is illustrated with the designed constructs C1-C15. B: SDS-PAGE gel results of the small scale expression and purification screens of the 15 CARD9 constructs, with the induced cell lysate on the left, followed by the proteins eluted from the Ni-NTA-column on the right. The constructs have been numbered and listed in Table 5-2.

#	construct	MW [kDa]	pI	expressed	soluble
1	1-115	15.3	5.75	x	x
2	1-142	18.43	5.77	✓✓✓	✓✓✓
3	1-174	22.3	6.25	✓✓	✓✓✓
4	1-184	23.5	6.25	✓✓	✓✓✓
5	1-200	25.3	6.32	✓✓✓	✓✓✓
6	1-223	28.0	6.08	✓✓✓	✓✓✓
7	1-231	29.0	6.43	✓✓✓	✓✓✓
8	1-243	30.5	6.87	✓✓✓	✓✓✓
9	1-277	34.3	6.33	✓✓✓	✓✓
10	1-297	36.9	6.04	✓✓✓	✓
11	1-336	41.6	6.26	✓✓✓	x
12	1-387	47.6	6.25	✓✓	✓
13	1-403	49.5	5.99	✓✓✓	x
14	1-420	51.4	6.08	✓✓✓	x
15	1-431	52.6	5.85	✓✓	x

Table 5-4 CARD9-CARD-CC constructs. 15 constructs of CARD9, including the CARD domain and additional amino acids on the C-terminal domain, have been screened by small scale expression and purification experiments. Results are illustrated as heavily ✓✓✓, moderately ✓✓ and minimally ✓ soluble constructs. Unexpressed/insoluble constructs are marked with x.

5.3.2.2 SEC multi-angle-laser-light scattering

To investigate the influence of the different coiled-coil fractions on protein oligomerisation, we used SEC-MALLS to investigate the molecular masses in solution of selected constructs, equivalent to the experiments on CARMA1 in section 5.2.2.2. The constructs chosen for these experiments were CARD9 aa 1-98, 1-142, 1-200 and 1-277. They have been successfully purified following the optimised protocol for CARD9 as described above in 5.3.1.3. The purified proteins were concentrated to 1.5 mg/mL and 3 mg/mL (precipitation was observed at higher concentrations) before loading onto a S200 10/300 GL column. For the shortest construct CARD9 aa 1-98, which only contains the CARD motif, SEC-MALLS analysis resulted in a main peak with a recorded molecular weight of 13 kDa consistent with a monomer (calculated $M_w = 13.4$ kDa). A small amount of the protein (<6 %) forms a dimer, showing up as a second peak with a

molecular weight of 26.2 kDa (Figure 5-15). For CARD9 aa 1-142, 1-200 and 1-277 however, analysis resulting in molecular weights corresponding to dimers with 34.7 kDa, 47.9 kDa and 68.4 kDa respectively. A small amount (<6%) of tetramers are formed, which seem independent of the protein concentration.

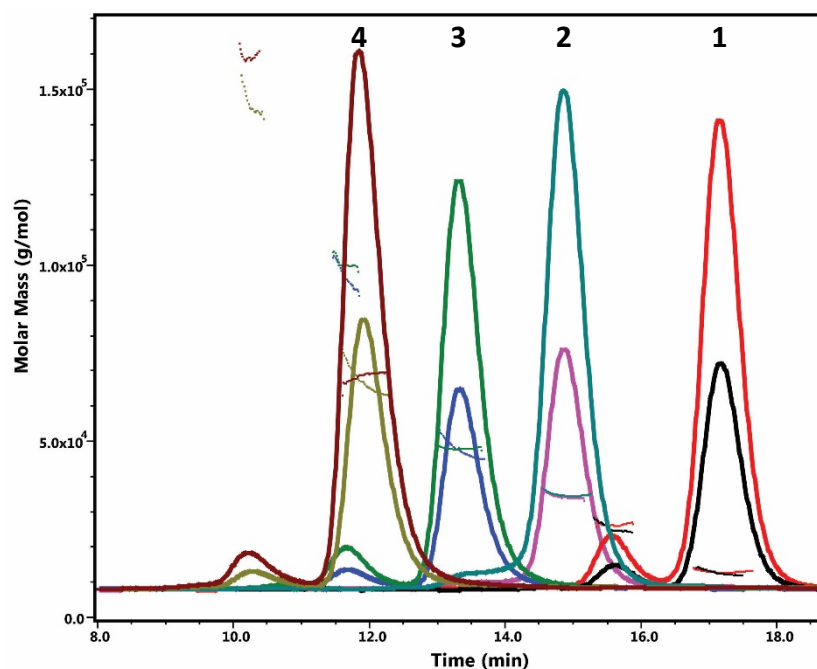


Figure 5-15 SEC-MALLS analysis of CARD9. The chromatogram shows the UV absorbance and the molecular weight calculation of the samples. The samples were loaded at 1.5 and 3 mg/mL. 1; SEC-MALLS analysis of CARD9 aa 1-98 resulted in a main peak with a recorded molecular weight of 13 kDa consistent with a monomer (calculated $M_w = 13.4$ kDa). 2: A molecular weight of 26.2 was recorded for CARD9 aa 1-142, corresponding to a dimer. 3 and 4: CARD9 aa 1-200 and 1-277 both are forming dimers with a molecular weight of 47.9 kDa and 68.4 kDa respectively, whereas <6% of the protein samples form a tetramer at both concentrations.

5.4 BCL10

5.4.1 Constructs, expression and purification

As it has been shown that Bcl10 forms a homotypic CARD interaction with either CARMA1 or with CARD9 in a cellular dependent context, successful purification of BCL10 would allow us to investigating these CARD-CARD interactions. However, expression tests of the CARD of BCL10 (BCL10-CARD aa 10-104) on a 750 mL scale showed that although this protein is expressed very well, it is mostly expressed as inclusion bodies (Figure 5-16). Therefore, a refolding protocol was developed for BCL10-CARD (aa 10-104), based on a protocol by Jang and Park (2011) [160]. Whereas standard refolding protocols include highly denaturing conditions such as 8M urea, this specific protocol contains just 3M urea in the lysis buffer, to partially unfold the proteins, assuming they form soluble aggregates. After centrifugation of the cell lysate, the supernatant was applied to a Ni-NTA column and the protein eluted by decreasing the pH. Refolding of BCL10 occurred by dialysis to remove the urea slowly. In parallel, a standard refolding protocol was performed as well. It differed from previous protocols by using 6 M guanidinium hydrochloride (GdmCl), several pellet washing steps, and included a fast protein-refolding on column. The refolding was performed by reducing the GdmCl concentration to 1 M and 0 M, before the protein was eluted with imidazole. Unfortunately, in both cases, the protein did not refold properly and aggregated.

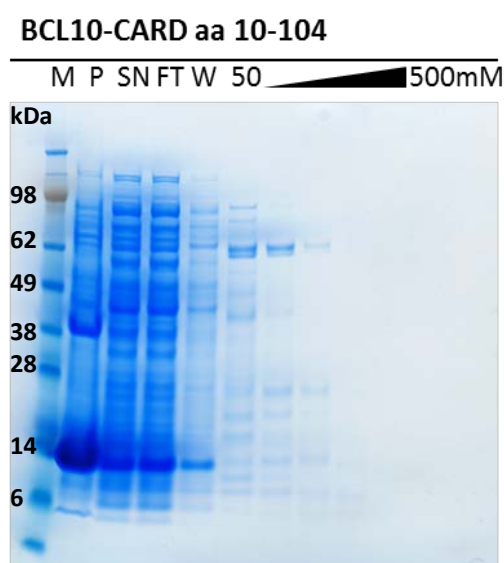


Figure 5-16 His₆-BCL10-CARD purification. SDS-PAGE analysis following expression of His-tagged BCL10-CARD aa10-104 (13.2 kDa). M; protein marker, P; pellet, SN; supernatant, FT; flow through, W; washing step. The protein should have been eluted by starting with an imidazole concentration of 75 mM and increasing to 500 mM.

Given these results, a different approach had to be taken. The structure of MAVS-CARD has been solved by Potter *et al.* in 2008 by x-ray crystallography as a MBP-fusion protein. Additionally, discussion with Dr Kovilen Sawmynaden in the group revealed, that he managed to purify a RIP2-MBP fusion protein successfully, whereas His₆-RIP2 and GST-RIP2 were insoluble. Therefore, fusion to MBP seemed a promising method for generating soluble BCL10 molecules. The pET-XM1 was used to generate several MBP-fusion proteins: MBP-BCL10 aa 10-107, aa 10-163 and aa 10-210. Upon further analysis, it became quickly clear that these constructs are soluble, but the proteins were prone to degradation making impossible to perform a full purification protocol. Looking again at the secondary structure prediction, the N-terminus was slightly extended and the construct MBP-BCL10 aa 6-123 was designed. The new

construct was expressed and purified following the standard protocol for MBP-constructs as described in section 2.3.8. Unlike previous BCL10 constructs, the MBP-fusion is not only soluble, but also less prone to degradation and can be purified obtaining a reasonable yield for further experiments. The purification steps were analysed via SDS-PAGE which is shown in Figure 5-17.

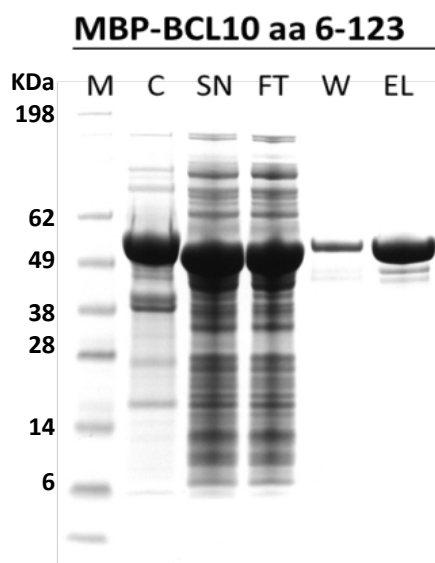


Figure 5-17 MBP-BCL10 aa 6-123 purification. SDS-PAGE analysis following expression of MBP-fusion BCL10 aa10-104 (54 kDa) using amylose resins. The protein was eluted with 10 mM Maltose. M; protein marker, C; induced cells, SN; supernatant, FT; flow through, W; washing step, EL; elution.

5.4.2 Biophysical analysis of MBP-BCL10

MBP, which has a molecular weight of 40 kDa, is a highly soluble protein and is nowadays often used as a solubility enhancement tag for fusion proteins. Although we obtained soluble MBP-BCL10 aa 6-123, it had to be considered that the actual BCL10 CARD-construct with only 16 kDa is much smaller than its fusion partner and the question remained if the CARD is in a folded state. For this reason, CD was used to analyse the secondary structure content of MBP-BCL10 aa 6-123 in comparison to MBP

alone. Spectra were recorded at 20 °C using a concentration of 0.15 mg/ml in 300 mM NaCl, 50 mM HEPES pH8, 1 mM EDTA, 1 mM TCEP and 10 mM Maltose. Figure 5-18, where data has been normalised to mean residue ellipticity, shows a high helical content for both MBP (52.5%) and MBP-BCL10 (47.9%). When these values are then applied to the total number of residues of each protein with MBP: 371 and MBP-BCL10: 489, we calculated a difference of 40 residues more alpha helical for MBP-BCL10, which correlates well with the 6 α -helices of the CARD domain fold. These results suggest that MBP-BCL10 is folded.

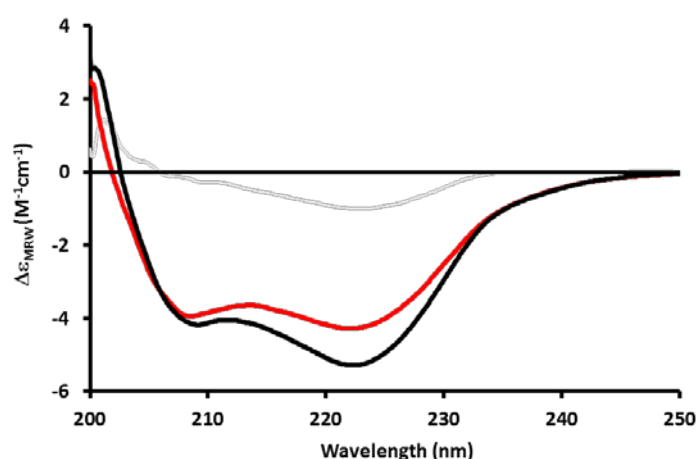


Figure 5-18 Far-UV CD spectra of MBP and MBP-BCL10 aa 6-123. Far-UV CD (200-250 nm) was carried out to analyse the secondary structure content of MBP (black) and MBP-BCL10 aa 6-123 (red). The difference of the two profiles is shown in light grey. Protein concentration was 0.15 mg/mL and spectra recorded at 20 °C.

In addition to the CD analysis, we wanted to investigate the oligomeric state and the homogeneity of the MBP-BCL10 aa 6-123 sample. The homogeneity of a protein sample can be just as important as purity regarding the likelihood to form a crystal lattice when setting up in crystallisation trays. The SEC-MALLS (S200 10/300 GL) was performed in 200 mM NaCl, 50 mM HEPES pH8, 1 mM EDTA, 1 mM TCEP and 3 mM

Azide at 4 °C with a loaded sample concentration of 2.5 mg/ml. As a comparison, a sample of the MBP-tag only was loaded. SEC-MALLS analysis of MBP-BCL10 resulted in one peak with an observed molecular weight of 53.3 kDa (Figure 5-19) in agreement with the calculated molecular weight of 54 kDa for a monomer. Analysis of the MBP sample provided a molecular weight of 38.9 kDa.

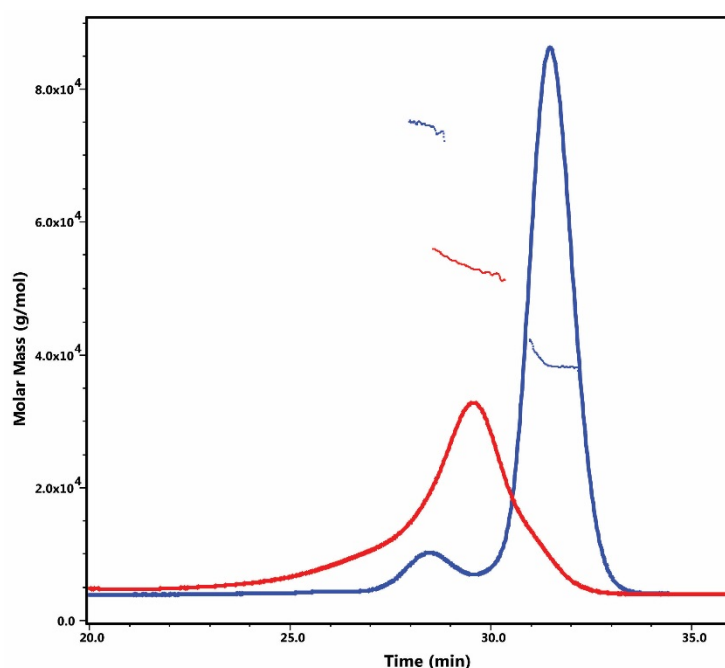


Figure 5-19 SEC-MALS analysis of MBP-BCL10 aa 6-123. The chromatogram shows the UV absorbance profile and the molecular weight calculation of the sample. Red: SEC-MALS analysis (S200 10/300 GL) of MBP-BCL10 aa 6-123 shows a single peak with a determined molecular weight of 53 kDa. The sample was injected at 2.5 mg/mL (47 μ M). Blue: The same experiment was carried out with MBP only, which resulted in an observed molecular weight of 38.8 kDa.

Having obtained several CARMA1 and CARD9 constructs as well as MBP-BCL10, binding studies and setting up crystallisation screens would have been the next step. However, at this point, Qiao *et al.* (2013) [103] published the filament assembly of BCL10 by

fluorescence polarisation assays, which is significantly enhanced in the presence of CARMA1. This study also involved the EM analysis of the BCL10-filament. These experiments were based on similar constructs as prepared already in this study: CARMA1-CARD with the addition of different coiled-coil segments and BCL10 as an MBP fusion protein. Taking into account our laboratory set-up and possibilities, this study changed our approach significantly: The focus was on looking further into the questions of CARD-filaments via fluorescence polarisation assays.

5.5 Fluorescence polarisation assays with BCL10 and CARMA1

It was recently shown that the signalosome formed by CARMA1, BCL10 and MALT1 is a helical filamentous assembly with CARMA1 nucleating BCL10 filament formation (Qiao *et al.* 2013). This study assessed the polymerisation quantitatively via fluorescence polarisation assays. Additionally, these assays showed that CARMA1 can enhance polymerisation of BCL10. In these experiments, polymerisation was initiated by cleaving BCL10 from its MBP-fusion protein.

Having also previously observed that MBP-BCL10 constructs are indeed soluble and folded, our goal was to set up a fluorescence polarisation assay to observe filament assembly between CARD9 and BCL10 using the published CARMA1-induced BCL10 polymerisation enhancement as a positive control. Having several CARMA1 and CARD9 constructs cloned and purified comprising different coiled-coil fragments, both CARMA1 and CARD9 could be compared regarding their ability to enhance BCL10 polarisation and the role of their coiled-coils.

5.5.1 Constructs, expression and purification

The MBP-BCL10 construct was re-cloned using the LIC-vector pHMBP, which has been described before with RIP2 (see section 4.5.1). This plasmid generates the desired protein construct including an N-terminal His₆-MBP-tag with a 3C cleavage site for tag-removal. Additionally, we chose to work with full length BCL10 for two reasons: firstly, to maintain consistency with the published data. Secondly, the BCL10 needs to be labelled with a fluorophore-maleimide at a location where the probe does not sterically interfere with any potential CARD-CARD interface. Hence, attaching the probe on the CARD itself must be avoided. Therefore, the two cysteines C29 and C57, which are located in the CARD were mutated to alanine using site-directed mutagenesis. Three further cysteines C-terminal of the CARD (C119, C121 and C215) still remain in the BCL10 full length protein. The newly generated full length His₆-MBP-(3C)-BCL10 C29/57A was expressed in 8L of LB media and purified via Ni-NTA using a lysis buffer of 500 mM NaCl, 50 mM Tris-HCl pH 7.5, 20 mM Imidazole, 0.5 mM TCEP and 10% glycerol. The different purification steps were analysed via SDS-PAGE and are shown in Figure 5-20. As an additional purification step as well as buffer exchange in preparation for the fluorophore labelling, the eluted protein was further purified on a S200 16/60 gel filtration column equilibrated in 500 mM NaCl, 50 mM Tris-HCl pH 7.5, 0.5 mM TCEP. The elution profile together with SDS-PAGE gel of the peak fractions can be seen in Figure 5-21.

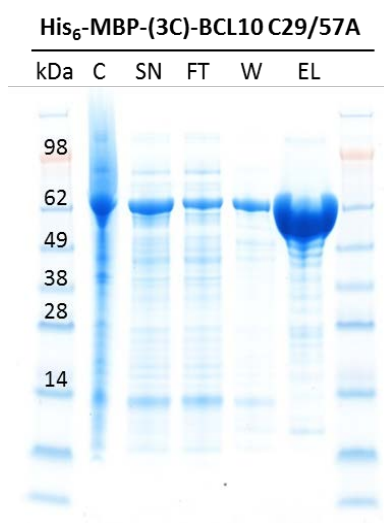


Figure 5-20 His₆-MBP-(3C)-BCL10 C29/57A purification. SDS-PAGE analysis following expression of His-MBP-tagged BCL10 full length double mutant (69 kDa). C; induced cells, SN; supernatant, FT; flow through, W; washing step, EL; elution. The protein was eluted with an imidazole concentration 500 mM.

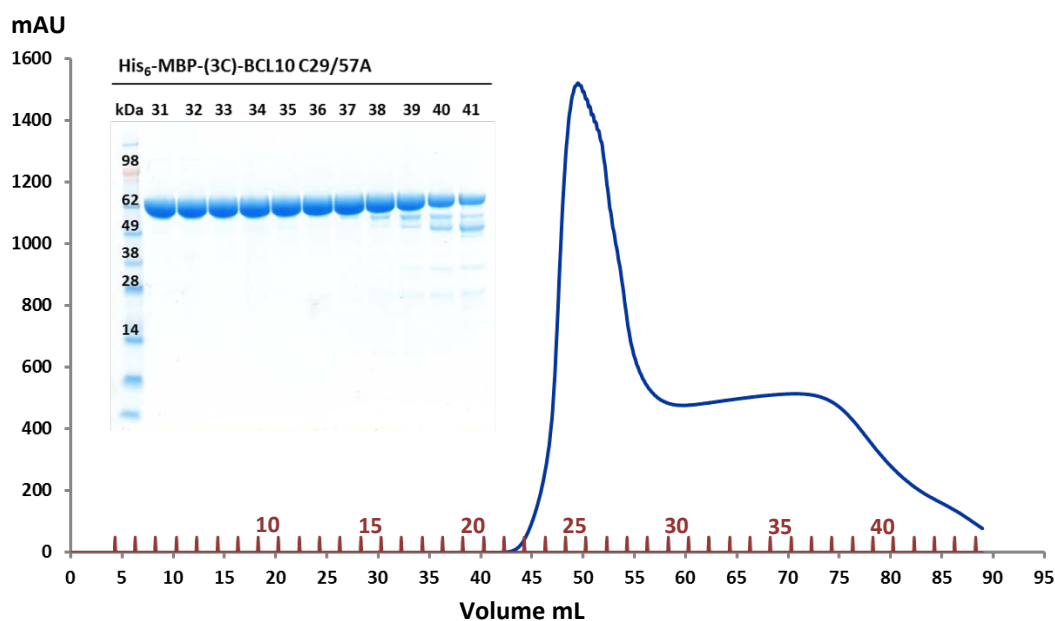


Figure 5-21 His₆-MBP-(3C)-BCL10 C29/57A gel filtration purification. Elution profiles of the His₆-MBP-(3C)-BCL10 C29/57A from a Superdex 200 16/60 gel filtration column together with the SDS-PAGE gel of the peak fractions. Fractions 34-37 were collected. The buffer used was 500 mM NaCl, 50 mM Tris-HCl pH 7.5, 0.5 mM TCEP.

5.5.2 Fluorophore labelling

The fluorophore used in this study was Alexa Fluor® 488 C₅ maleimide. The protein sample was concentrated to 50 µM prior mixing with a 3-fold molar excess of Alexa Fluor® 488 C₅ maleimide. After 2h incubation at room temperature, another gel filtration chromatography was performed using 150 mM NaCl, 20 mM Tris-HCl pH 7.5 and 0.5 mM TCEP to remove free dyes. At this step, around 80% of the protein was in the void, which can be seen in the gel filtration profile of Figure 5-22. The following monomeric peak containing the labelled BCL10 was collected and concentrated to 10 µM for further studies.

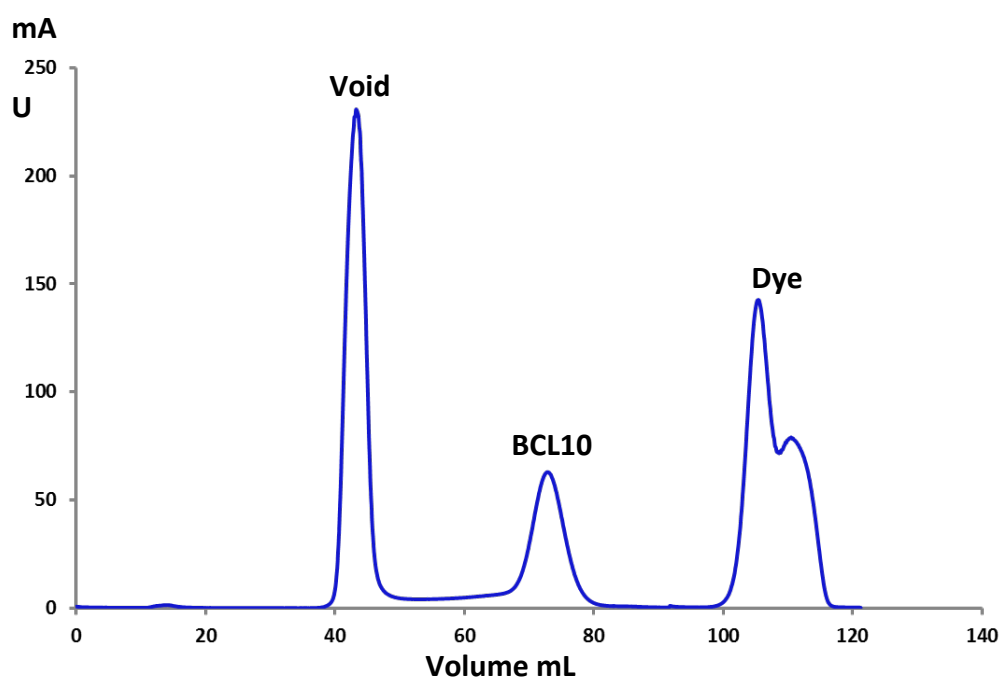


Figure 5-22 His₆-MBP-(3C)-BCL10 C29/57A gel filtration purification after labelling. Elution profiles of the Alexa Fluor® 488 C₅ labelled His₆-MBP-(3C)-BCL10 C29/57A from a Superdex 200 16/60 GL gel filtration column together. The peak of monomeric BCL10 was collected and pooled. The buffer used was 150 mM NaCl, 20 mM Tris-HCl pH 7.5, 0.5 mM TCEP.

5.5.3 Biophysical analysis

5.5.3.1 UV spectrum

To estimate the protein concentration as well as to determine the degree of labelling, a UV spectrum was recorded to measure the absorbance of the conjugate at 280 nm and 495 nm. The concentration for the labelled protein was calculated as described in chapter 4.5.3.1. The UV spectrum is shown in Figure 5-23. An absorption of 0.419 at 495 nm and 0.284 at 280 nm was measured. With an extinction coefficient of $72,000 \text{ cm}^{-1}\text{M}^{-1}$ and $71,000 \text{ cm}^{-1}\text{M}^{-1}$ for Alexa Fluor® 488 C₅ and MBP-BCL10 respectively, a protein concentration of $10 \text{ }\mu\text{M}$ and a dye concentration of $19 \text{ }\mu\text{M}$ were calculated. There are two possibilities to explain this discrepancy: considering the BCL10 construct contains three cysteines, one could conclude that 63% of the cysteines are labelled. On the other side, considering the size of the fluorophore, steric effects could mean that only one label can possibly be attached to either C119 or C121. This would mean 95% of the protein sample is labelled, a similar result as achieved with RIP2 in Chapter 4.

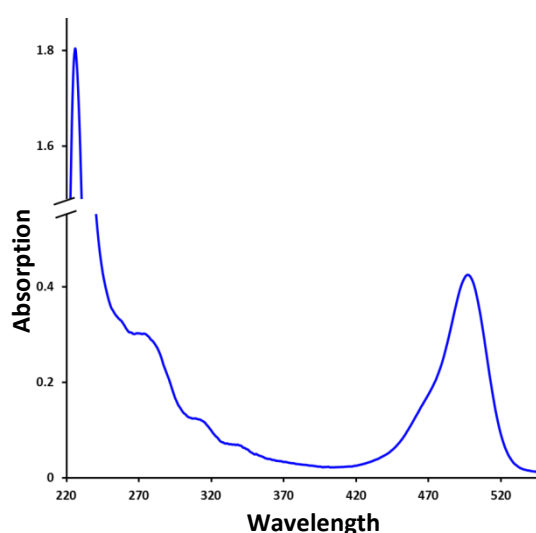


Figure 5-23 UV spectra of the labelled His₆-MBP-(3C)-BCL10 C29/57A. Protein concentration was calculated by recording a UV spectrum and measuring the absorption at 280 nm (protein +dye) and 495 nm (dye).

5.5.3.2 CD spectroscopy

To ensure this construct remained folded during the labelling process, the sample of the labelled MBP-BCL10 was used for CD spectroscopy to determine the secondary structure content. The MBP-BCL10 construct is clearly alpha-helical (Figure 5.24). The intensity of this spectrum, however, is probably affected by possible errors in the estimate of protein concentration due to the fluorophore (see 4.4.3.1)

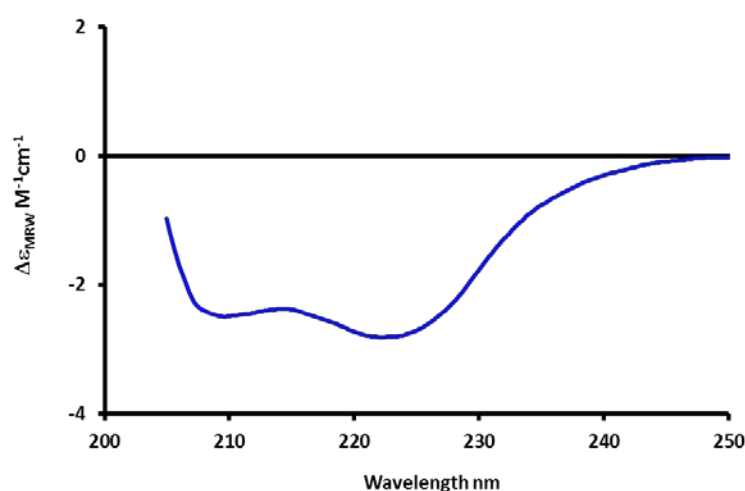


Figure 5-24 Far-UV CD spectra of labelled His₆-MBP-(3C)-BCL10 C29/57A. Far-UV CD (220-250 nm) was carried out to analyse the secondary structure content of the labelled His₆-MBP-(3C)-BCL10 C29/57A. Protein concentration was 0.15 mg/mL and spectra recorded at 20 °C.

5.5.4 Fluorescence polarisation assay

In preparation for the fluorescence polarisation assay, the excitation and emission spectra of the Alexa Fluor® 488 labelled MBP-BCL10 were recorded (Figure 5-25). Both excitation and emission maxima are close together an excitation of 492 nm and emission of 522 nm was used for the assay. Fluorescence polarization assay was performed at 25 °C in a buffer containing 150 mM NaCl, 20 mM Tris-HCl pH 7.5 and 0.5

mM TCEP. The sample volume was 80 μ L with a MBP-BCL10 concentration of 10 μ M. About 100 seconds after starting the experiment with a recording time of 2h (7200 seconds) 5 μ g of GST-3C protease (purified in-house) was added to the cuvette for removal of the MBP-fusion to allow BCL10 polymerization. To investigate the stimulation of BCL10 polymerisation by CARMA1 the experiment was repeated in the presence of 2 μ M (end concentration) CARMA1 aa 18-200 in the CARMA1-induced BCL10 polymerization experiment.

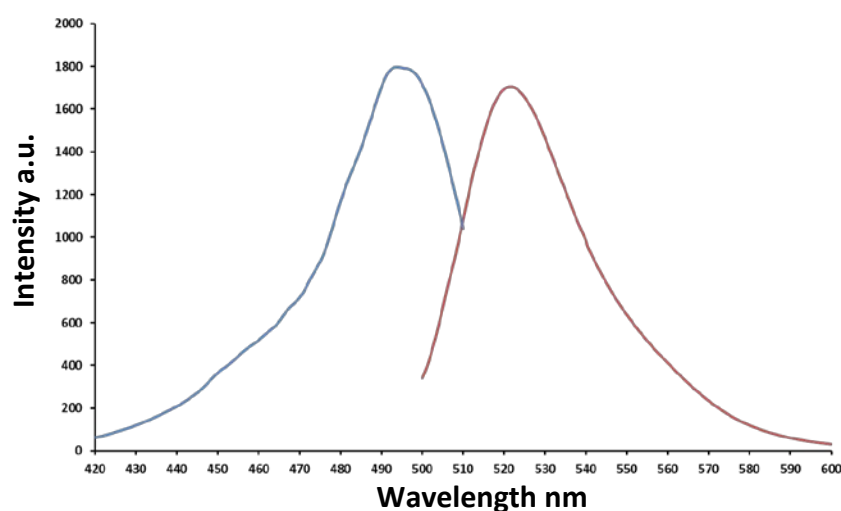


Figure 5-25 Fluorescence spectra. Excitation (blue) and emission (red) spectra of Alexa® Fluor 488 labelled MBP-BCL10 C29A/C57A.

As in the FPA experiments in Chapter 4, the 3C cleavage here resulted in a decrease in fluorescence polarisation as the molecular weight of the labelled protein drops from 69 to 26 kDa. However, the decrease is much smaller (130 mP to 120 mP) compared to the decrease seen upon MBP-RIP2 cleavage (220 mP to 90 mP). This is mainly due to a much lower initial value of the polarization. Since the molecular weight of MBP-BCL10 is even higher than that of MBP-RIP2 (69 vs 55 kDa) the difference is likely due to the fact that the Alexa fluorophore is much more flexible when attached to BCL10

compared to RIP2 and therefore, its tumbling is not strongly restricted by the size of the protein. This is in line with the fact that the Alexa-488 fluorophores are coupled to the cysteine residues on the unstructured C-terminus of BCL10 whereas RIP2 has the Alexa-488 fluorophore attached only on a 16-residues extension preceding the CARD. There is no further significant change in polarisation after the initial drop over the time scale of the experiment either in the presence or absence of CARMA1 molecules (Figure 5-26), so BCL10 does not polymerize under our experimental conditions.

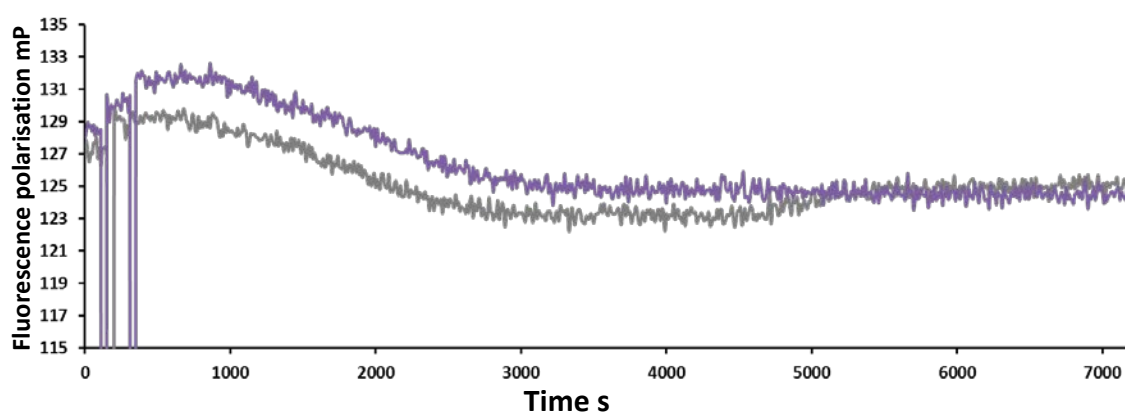


Figure 5-26 Fluorescence polarization assay. Alexa 488-labeled MBP-BCL10 C29/57A in the presence (purple) and absence (grey) of substoichiometric amounts (1/5) of CARMA1 aa 18-200 upon MBP removal by 3C protease cleavage. The temporary fluorescence drops at the beginning of the experiment represent the addition of CARMA1 and/or 3C protease.

As no polymerisation could be detected with this sensitive assay in the contrast to published data [103], it was considered that the added glycerol (7 μ L of 50% glycerol) from the 3C protease stock may play a role in suppressing oligomerisation. For this reason, another fluorescence polarisation assay was recorded where glycerol from the 3C stock solution was removed by dialysis prior the experiment. Figure 5-27 shows that

the glycerol was not inhibiting any polymerisation as no change in fluorescence polarisation was detected.

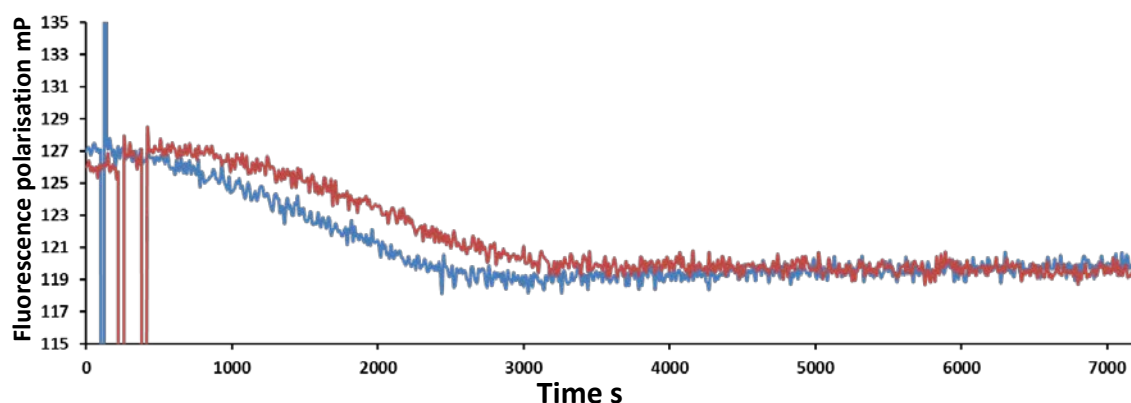


Figure 5-27 Fluorescence polarization assay without glycerol. Alexa 488-labeled MBP-BCL10 C29/57A in the presence (red) and absence (blue) of substoichiometric amounts (1/5) of CARMA1 aa18–200 upon MBP removal by 3C protease cleavage. Any glycerol from the 3C stock solution has been removed via dialysis prior to the experiment.

In both assays recorded, we could clearly monitor that the GST-3C protease requires about 2000s (~30min) to fully cleave all MBP-BCL10. To rule out every factor and possibility in order to know the reason why the fluorescence polarisation assay doesn't show any BCL10 molecule polymerisation, a third attempt was carried out using a different 3C protease: an in-house purified His₆-3C protease. The new assay carried out, which can be seen in Figure 5-28, shows the cleavage time was decreased to about 750s (12 min). However, no difference in BCL10-polymerisation compared to the previous assays could be monitored. After the experiments, the samples were analysed via SDS-PAGE. According to the gel, all MBP-BCL10 has been cleaved during the experiment (Figure 5-29).

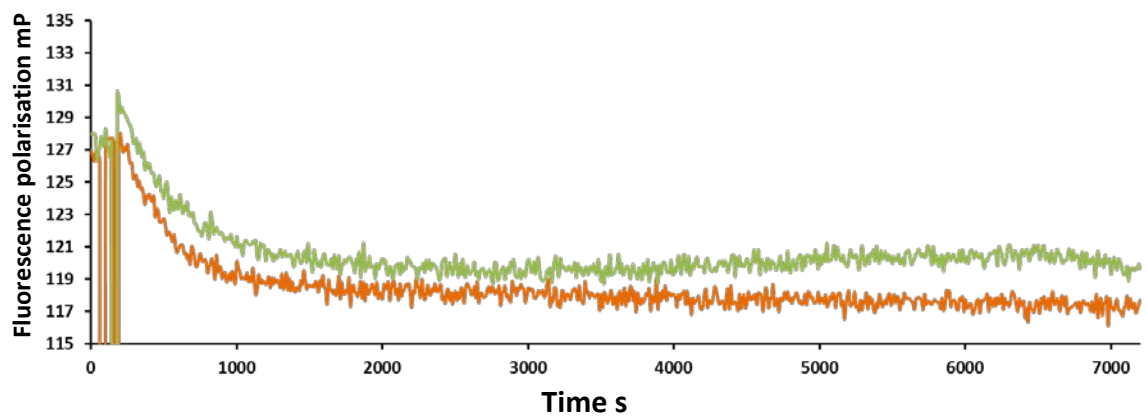


Figure 5-28 Fluorescence polarization assay with His₆-3C. Alexa 488-labeled MBP-BCL10 C29/57A in the presence (orange) and absence (green) of substoichiometric amounts (1/5) of CARMA1 aa18–200 upon MBP removal by 3C protease cleavage. In this particular case, we used the in-house purified His₆-3C protease.

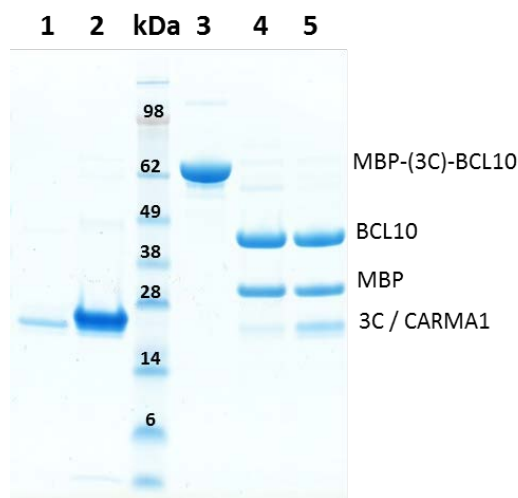


Figure 5-29 SDS-PAGE analysis post FPA. 1: 3C protease (21.7kDa), 2: CARMA1 aa 18-200 (24 kDa), 3: uncleaved MBP-BCL10, 4: post FPA sample without CARMA1, 5: post FPA sample with CARMA1.

5.6 Summary

In this chapter, the focus was set on CARMA1 and CARD9, both of which are adaptor proteins that interact with BCL10 *via* a CARD-CARD interaction, an essential step to initiate the formation of a signalosome involving mucosa-associated lymphoid tissue MALT1 to mediate NF- κ B activation. CARMA1-CARD is a well expressed protein construct. No crystals were obtained, although analysis revealed its potential to investigate the structure via NMR spectroscopy. At this point, the crystal structure of CARMA1-CARD was published.

CARD9-CARD on the other hand needed further investigations *via* thermofluor before a suitable condition for purification was finally found. Initial crystallisation experiments were set up before the focus was then extended to CARMA1 and CARD9 constructs involving segments of the following coiled-coil domain. SEC-MALLS analysis with constructs containing up to 2 of the 4 coiled-coil segments showed that the CARDS alone exist as monomers, while the coiled-coil drives the molecule into dimers. However, neither CARMA1 nor CARD9 constructs containing the full 4 segments of the coiled-coil domain could be purified successfully.

BCL10 on the other hand, is a protein, which is poorly soluble and after failed several purification protocols, including refolding experiments, we managed to purify BCL10-CARD as an MBP fusion protein. Recent studies showing BCL10 forming filaments was addressed by getting into contact with the laboratory of Professor Hao Wu and setting up similar fluorescence polarisation assays. However, no polymerisation of BCL10 CARDS could be monitored upon MBP-tag removal, neither in absence or presence of CARMA1.

Chapter 6

Discussion and outlook

6.1 Overview

This thesis was aimed at providing molecular insight into the mechanism by which CARDS mediate the formation of multi-protein assemblies. Experiments were initially designed assuming that CARDS may either form 1:1 complexes or may require an additional oligomerisation domain for complex formation, as observed in the apoptosome structure.

Another model assumed, that CARDS may be able to form individual signalling platforms as observed for members of the death domain subfamily. The focus was set on finding a general mode of CARD-CARD interactions by investigating these possibilities. However, during the course of this study, the CARD-containing proteins have been suggested to form filaments, which established a complete new principle regarding the binding characteristics of CARDS.

In this Chapter, I will discuss our findings regarding the intramolecular interaction of NOD2 tandem CARDS from mutational analysis involving mainly ITC, as well as the intermolecular interaction with RIP2 gained from pull-down experiments. This is followed by a short look into the purification and characterisation of different CARMA1 and CARD9 constructs. To address the latest behavioural insight gained during the time course of this study regarding CARDS as filament forming motifs, fluorescence polarisation assays will be discussed in the last part of this section.

6.2 CARDS - expression and solubility screens

A main reason for the lack of structural information on CARDS is likely their tendency to aggregate when they are purified as individual domains. The first aim was therefore to screen all human CARDS in co-expression and pull-down experiments with the goal of finding soluble CARD-pairs, which can be taken further for biochemical, biophysical and structural investigations. These experiments were designed hoping that co-expression would increase the solubility of individual CARDS, but more importantly, we hypothesised that CARDS bound to the beads during pull-down experiments may potentially mimic an additional oligomerisation domain should this be required for a CARD-CARD interaction. In this case, the beads would bring CARDS into close proximity to form the platform needed to pull down a binding partner.

LIC cloning turned out to be a straightforward method to create all CARD constructs for a genome wide screen. The newly generated CARD constructs were first tested for their expression and solubility individually. Some variables were set after initial experiments (e.g induction temperature was set to 18 °C) whereas others were based on previous experience in the group (e.g. *E.coli* strain: BL21 (DE3) Gold). Comparing both His₆- and GST-constructs, the number of soluble constructs as His₆-proteins was significantly higher. This was unexpected because the 26 kDa GST-tag can enhance solubility by stabilising the fusion protein [147, 161]. However, in this case it could be that the dimerisation of the GST did enhance precipitation through aggregation. The observed differences in the expression and solubility profiles between all CARDS tested remains unclear at this stage. Similarly, we do not understand at this point why the newly designed pGST vector showed such low protein expression. This is likely to be a problem with the new vector itself and would need to

be tested with established proteins for which we know the amount of expression and solubility.

The high amount of insoluble constructs found in these two screens is not entirely surprising, as many CARDS described in the literature seem to have low solubility. The tendency to aggregate is more and more being related to their biological function with forming homotypic protein-protein interactions to mediate higher-order complex formation and filamentous assemblies regulating downstream signalling. Additionally, a number of CARD-containing proteins have been reported to remain in an auto-inhibited state until activation occurs via phosphorylation. In these cases, the CARDS are undergoing intramolecular interactions and are inaccessible when inactive, hence, hydrophobic residues are buried and therefore the protein is monomeric. These factors could explain why the purification of isolated domains may interfere with their natural environment. As seen in Table 1-2, proteins containing CARDS belong to different families including receptor, scaffold, effector or adaptor proteins. The question arises if there is any link between the solubility and the identity of the protein. The 14 His₆-CARDS which turned out to be partially or completely soluble, these CARD proteins are listed together with their overall domain structure, followed by their functionality in Table 6-1. However, there seem to be no obvious commonality visible between the CARDS from multi-domain receptor proteins as well as single domain adaptor proteins, which have turned out soluble. With 16 out of 33 isolated CARD structure solved at this point, it is not surprising that 11 are here listed with the soluble constructs obtained.

CARDs	Domain structure	Function	PDB
NOD1 ✓	NBD-CARD	Receptor	2NZ7
ASC ✓	Bipartite-CARD	Adaptor	2KN6
APAF1 ✓✓✓	NBD-CARD	Receptor	3YGS
NALP1 ✓✓✓	NBD-CARD	Receptor	3KAT
Caspase 9 ✓	Bipartite-CARD	Initiation / Effector	3YGS
CARD8 ✓	Bipartite-CARD	Inhibitor	4IKM
CARD9 ✓✓	Coiled-coil	Regulator	-
RAIDD ✓✓✓	Bipartite-CARD	Adaptor	3CRD
Iceberg ✓✓	Shorthair-CARD	Inhibitor	1DGN
RIG-I ab ✓✓	Multipartite-CARD	Receptor	4P4H
MDA5 b ✓	Multipartite-CARD	Receptor	-
CARMA1 ✓✓✓	Coiled-coil	Regulator	4LWD
CARD6 ✓	Longhair-CARD	Regulator	-
ARC ✓	Shorthair-CARD	Inhibitor	4UZO

Table 6-1 Soluble His₆-CARD constructs and their protein function. Listed are the His₆-CARD constructs, which were either soluble or only partially soluble together with their domain structure and protein function. Results are illustrated as heavily ✓✓✓, moderately ✓✓ and minimally ✓ soluble constructs.

Regarding the disappointing expression level of the GST-tagged CARD constructs using the newly designed vector pGST, a complete re-cloning of all human CARDs into a new GST-vector for co-expression and pull-down screens was not taken any further because Qiao *et al.* (2013) [103] published the filament formation of BCL10-CARD at this point. This completely changed the aspect of how to further approach the investigation of CARDs if the idea of filamentous CARD assembly wanted to be addressed in the time course of this work. For this reason, the attempt to perform a wide range of co-expressions and pull-down experiments was dismissed and the focus was set on a selection of CARDs: Work with CARMA1, CARD9 and BCL10 has

already been performed and provided a good system to look further into the polymerisation of CARDS. Another focus was set on the tandem CARDS of NOD2 as preliminary work had already been performed in our group and especially because the structure of the N-terminal NOD2-CARDa was recently solved by a former group member. This provided the opportunity to compare intra and intermolecular CARD-CARD interactions.

6.3 Intramolecular interaction between the tandem CARDS of NOD2

NOD1, NOD2 and other members of the NLR protein family have emerged as key cytosolic pattern recognition receptors for the detection of pathogens and endogenous danger signals. Stimulation of these receptors activates signalling cascades that lead to NF- κ B activation and subsequently to transcription of genes mediating pro-inflammatory immune responses. One of the main differences between NOD1 and NOD2 is that NOD2 contains two CARDS in tandem. The question immediately arises, if this second CARD copy is biologically relevant and if so, what difference in function is between the two CARDS of NOD2? An intramolecular interaction between CARDa and CARDb had been suggested initially by yeast two-hybrid assays by Wagner *et al.* (2007) [130] and has later been confirmed on the purified protein level by our group [131]. At that point, only one CARD-CARD complex structure was available, which was that of APAF1-CARD interacting with Procaspase-9-CARD *via* a type 1 interaction (3YGS, also Figure 1-12). In this 1:1 complex, strong charge-charge interactions play a crucial role. Fridh and Rittinger

analysed the intramolecular interaction of NOD2 CARDS by ITC, mainly focusing on a potential type 1 interaction between CARDa and CARDb by targeting acidic as well as basic residues in each CARD. Mutants E69K, E72K and R86A, all located in CARDa, were expressed and soluble, but ITC results showed that neither of the mutations had any effect on complex formation. This strongly suggests that CARDa and CARDb are not interacting via a type 1 interface. This suggestion is in line with the results of the NMR saturation transfer difference spectroscopy, where two tryptophans of CARDa seemed to be located in the interface.

My initial approach to characterise this intramolecular interaction of NOD2 in more detail was to carry out NMR titration experiments, using NOD2-CARDa tryptophan mutants to establish, which of the five potential tryptophans are at the interface with CARDb. However, recombinant protein expression of the NOD2 tandem CARDS was significantly reduced in minimal (M9) media and even using a commercial 'rich' isotopically-labelled medium it was not possible to produce sufficient protein to reach a satisfying signal-to-noise ratio when recording ^{15}N -HSQCs. This low protein expression suggested that the mutations changed the protein stability and structure severely. Nevertheless, NMR spectra of the proteins indicated that all the tryptophan mutants were folded and it is not clear at present why protein expression was so dramatically decreased in minimal medium.

As an alternative method to study the role of tryptophan residues in the intramolecular interaction, we have chosen ITC as used by Fridh and Rittinger. The ITC data obtained in this work supports the preliminary NMR data, as the intramolecular CARD-CARD interaction of NOD2 is lost when W59 and W63 are mutated to alanine.

Structural alignments with available tandem death-fold domains were carried out to test if an intramolecular interface containing W59 and W63 may be similar to other tandem motifs. A good overlay was achieved between NOD2-CARDa and the tandem DEDs of MC159 (Figure 4-7). Another available structure of a tandem motif is that of the tandem DED of Caspase-8 and interestingly, in both structures the intramolecular interface between DED1 and DED2 is mainly formed by the hydrophobic residues from helices $\alpha 2a$ and $\alpha 5a$ of DED1 and helices $\alpha 1b$ and $\alpha 4b$ of DED2. A sequence analysis is displayed in Appendix A 6-1. The tryptophans W59 and W63 of NOD2 are located at the same hydrophobic DED1 patch forming the interface with DED2. Based on this idea, other NOD2 mutants were analysed to test the idea of NOD2 interacting in a similar manner as tandem DEDs and to exclude a type 2 and type 3 interaction. Interestingly, another mutant was identified that was unable to bind CARDb: S49A. This residue, which lies in the loop between helices $\alpha 1$ and $\alpha 2$, was originally selected because it appeared to be located in a type 3 interface. All CARDa residues successfully purified and analysed by ITC are mapped onto CARDa, shown in Figure 6-1.

If W59, W63 and S49 were essential for binding, then the interface would span across the whole of helix $\alpha 2$ of CARDa. However, such an interface seems not fully accessible, because of the position of the elongated helix $\alpha 5$. Though one could assume that helix $\alpha 5$ may be shorter in the presence of CARDb, with the remaining residues forming helix $\alpha 6$, which, in its right place, is folded towards the N-terminal domain between helices $\alpha 1$ and $\alpha 5$ (Figure 6-1). This would open up the space spanning from tryptophans W59 and W63 till residue S49 thereby allowing binding of CARDb.

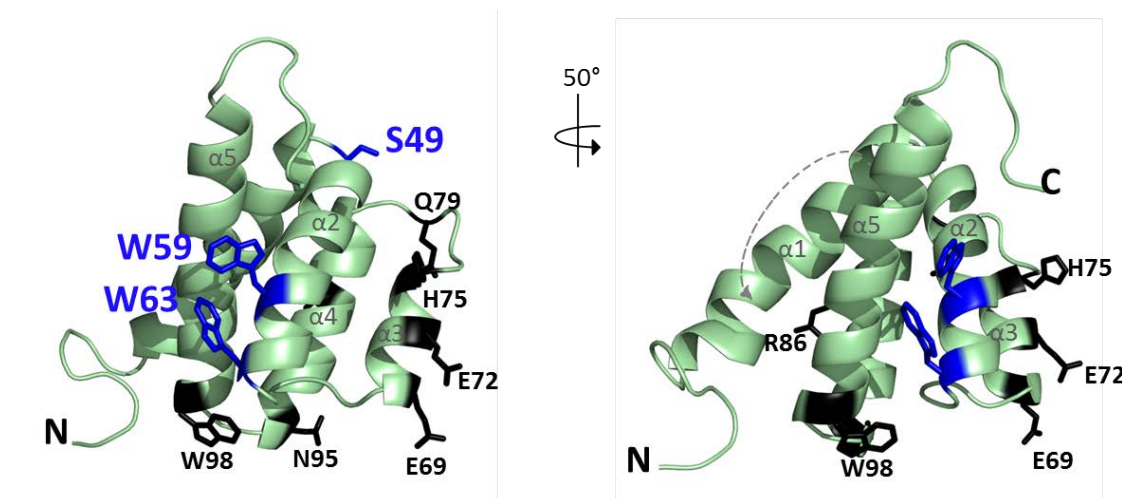


Figure 6-1 Residues analysed *via* ITC mapped onto NOD2-CARDa. The solution structure of CARDa is shown with S49, W59 and W63 highlighted in blue. Other residues tested without showing any effect on CARDb binding are shown in black. In presence of CARDb, helix $\alpha 5$ is believed to be shorter to form helix $\alpha 6$, which would be folded back (dashed arrow) towards $\alpha 1$ and $\alpha 5$, to complete the 6 alpha-helical bundle of a death-fold.

In order to determine residues in NOD2-CARDb essential for CARDa binding, a model of CARDb (Swiss Model) was calculated and then overlaid with DED2 of MC159. Selected mutants, which could be expressed and purified were D133A and H137A, with negative controls E158A and Q201A. All of these showed binding and no CARDb residue could be identified as being essential for the CARDa-CARDb interaction.

However, it has to be considered that we are using only alanine mutations in this study. Comparison with charge reversal mutations could give more insight into the interaction. For example, in a study of NOD1 and RIP2 binding *via* luciferase reporter assays, Boyle *et al.*[162] showed that the NOD1 mutant E56A had no effect on RIP2

binding, whereas a charge reversal mutation disrupted the interaction almost completely. Nevertheless, such experiments can also be misleading and the loss of binding is not a cause of specifically a charge reversal: Manon *et al.* [163] analysed the same interaction and showed that a charge reversal on a different residue, at NOD1 R69, was able to disrupt the interaction with RIP2. However, subsequent structural analysis of NOD1 showed that R69 forms an internal salt bridge with residue D73, therefore, the loss of binding is likely due to the destabilisation of the protein.

6.4 Intermolecular interaction of NOD2 and RIP2

To study the interaction between the tandem CARDS of NOD2 and RIP2, pull-down experiments were performed following a protocol developed in the group [131]. However, the GB1-RIP2-CARD was replaced by MBP-RIP2-CARD for consistency with the MBP-BCL10 studies described in Chapter 5.

Although RIP2 is the best studied interaction partner of NOD2, still many questions remain about the molecular details of their interaction. Furthermore, due to expression problems with individual CARDS, it is still unclear whether both CARDS are essential for RIP2 binding and subsequent activation of NF- κ B signalling. Some reports suggested that neither CARDa nor CARDb alone is able to bind RIP2 [64, 130] and point mutations in either CARDS have been shown by Tanabe *et al.* [164] to prevent NF- κ B signalling. On the other hand, the shortest construct reported so far able to bind RIP2 is NOD2-S (aa 1-180) designed by Rosenstiel *et al.* [165]. This construct contains CARDa and the first three helices of CARDb. Interestingly, even though NOD2-S is able to bind RIP2, overexpression does not activate NF- κ B unlike the overexpression of the

complete tandem CARDs. Probably the most contradictory studies attempting to characterise the interface between NOD2 and RIP2 involve the results of Wagner *et al.* [130] and Fridh and Rittinger: Wagner reported that the conserved acidic patch involving residues E69 and D70 in NOD2-CARDa are essential for RIP2 binding based on yeast-two hybrid assays, whereas Fridh and Rittinger concluded that E59K and E72K are not essential for RIP2 binding based on pull-down assays in *E.coli*. Instead they showed that basic residues R38 and R86 were important for recognition of conserved acidic residues on RIP2-CARD (D461 and D473). Such an interaction would correspond to a type 1 interface.

As always with mutational studies, one has to take into account that the loss of interaction can result from a mutation, which destabilises the fold rather than disrupting the binding. With NOD2-CARDa residues R38A and R86A so clearly losing their ability to bind RIP2-CARD, I purified the R38A and R86A MBP-fusion constructs and recorded 1D proton NMR spectra (data not shown). In support to Fridh and Rittinger's study, peaks were clearly visible > 9 ppm and < 0.5 ppm and the spectra similar to the wild type. Therefore, we conclude these constructs are folded.

Having already designed a large number of NOD2 mutants for the ITC experiments to investigate intramolecular CARD-CARD interactions, the idea was to take these mutants further to study RIP2 binding by performing further pull-down experiments. R38A was repeated to serve as a negative control. Interestingly, all new mutants tested (CARDa: S49A, W59A, W63A, H75A, Q79A, N94, W98A. CARDb: D133A, H137A, E158A, Q201A) are able to bind RIP2-CARD. Mapping all residues tested via pull-down on the NOD2-CARDa structure (Figure 6-2) clearly shows that the basic patch including R38 and R86 is at the opposite end of the residues detected to be

involved in CARDb binding (S49, W59, W63). This would make sense as two different binding sites are expected for intra- and intermolecular interactions. Although no essential residues of CARDb could be detected so far to be important in RIP2 binding (Fridh and Rittinger tested CARDb mutants A106V and Q164K, whereas this study added D133A, H137A, E158 and Q201A), I personally doubt that NOD2 has an arbitrary duplicate CARD and suspect that they both have different functions and stabilise each other.

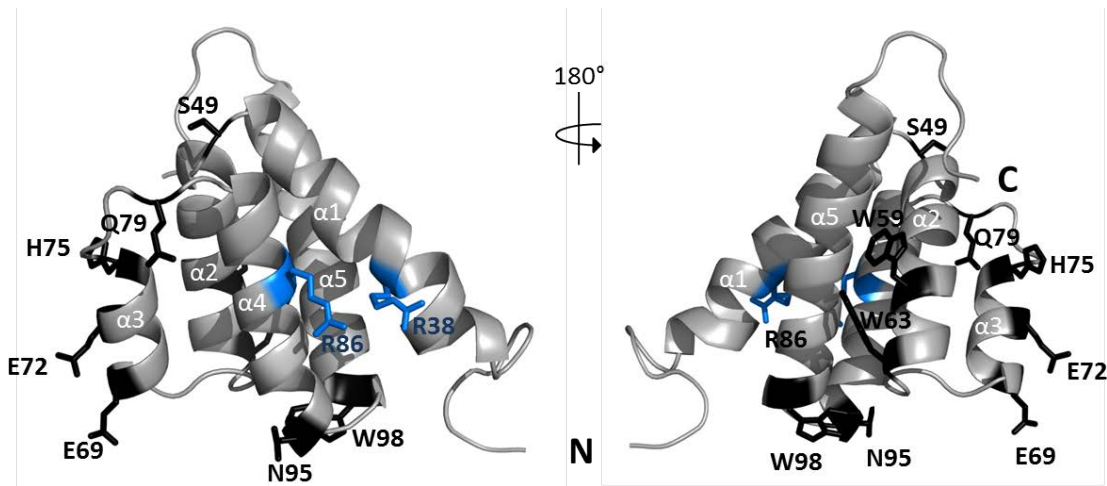


Figure 6-2 NOD-CARDa and RIP2 interaction. The mutational analysis of NOD2-CARDa and RIP2-CARD binding identified a basic patch formed by R38 and R86 (shown in blue) in NOD2-CARDa to be essential for RIP2 binding. The two arginines are located at the opposite end of the previously defined intramolecular binding site with CARDb.

6.5 CARMA1 and CARD9 - the CBM complex

The small scale expression and purification screen drew our attention towards an exceptionally well expressed and soluble CARD: CARMA1. CARMA1 is a scaffold protein and a key player in the formation of the so-called CBM-L signalosome. This complex comprises CARMA1, BCL10 and MALT1 and acts downstream of the antigen receptor in lymphocytes to activate NF- κ B. The CBM-L complex is essential for normal cellular functions. Malfunction in activity leads to inflammatory disorders, making it a viable target for novel therapeutic design. When we identified the CARD of CARMA1 as a highly soluble protein, no structural details of neither CARMA1 nor BCL10 were available, which made these two proteins especially interesting for me. Furthermore, when I decided to focus on the CARMA1-BCL10 interaction, I also included CARD9 in my study, as it has a similar domain structure to CARMA1 (CARD followed by a long coiled-coil) and also binds BCL10 to form the equivalent CBM complex in myeloid cells.

The purification of CARMA1-CARD was very straightforward and NMR analysis demonstrated that CARMA1-CARD was properly folded. Therefore, it was taken further to record initial 3D NMR experiments to test its suitability for further structural studies. In parallel, crystallisation trays were started. However, before structural analysis could be completed, Li *et al.* published the crystal structure of CARMA1-CARD, which shows a typical 6-helical fold characteristic of proteins of the death domain superfamily. The focus consequently was set on larger constructs that include the adjacent coiled-coil domain, as an oligomerisation domain might be required to form a platform to bind other CARDS.

Although the CARD9-CARD seemed partially soluble in our initial expression screens, large scale expression/purification yielded largely insoluble protein. However, modification of domain boundaries, coupled with ThermoFluor assays using commercial solubility and stability buffer screens, allowed us to identify better lysis buffer conditions containing LiCl and BIS-TRIS propane pH 7.5, which excitingly was very effective and led to the successful purification of several stable CARD9-CARD-coiled-coil constructs of different lengths.

To characterise the self-association of CARMA1 and CARD9, SEC-MALLS experiments were performed. SEC-MALLS analysis of various CARMA1 constructs showed that the CARD itself is a monomer in solution, while addition of the first of four coiled-coil fragments, induces dimerisation. Unfortunately, adding a further fragment to the protein so the construct contains half of the complete coiled coil (aa 18-299), shifted the protein to be eluted in the void peak, corresponding to a molecular weight > 1300 kDa. Several attempts to purify CARMA1-CARD together with the full coiled-coil were similarly unsuccessful. These data on CARMA1 are consistent with the later published observations by Qiao *et al.* (2013) [103], which showed by EM negative stain analysis that the construct aa 1-295 (eluted in the void) forms protein clusters.

CARD9 behaved overall similar to CARMA1, although SEC-MALLS analysis of the construct containing half of the coiled-coil domain was successful and demonstrated a dimeric protein. However, several attempts to purify the entire CARD-coiled-coil construct were not successful either, indicating that adding longer coiled-coil fragments is pulling CARD9 into aggregation as well. This suggests that not only domain structure and cellular function of CARMA1 and CARD9's are similar, but also the behaviour of their individual molecular building blocks.

6.6 Study of filamentous assemblies

BCL10 is a protein prone to aggregation. During the course of this work, other groups also reported their difficulties with BCL10 [156]. The use of an MBP-tag finally proved to be successful and several BCL10 constructs appeared to be soluble. However, during the course of this work it was shown that BCL10 forms a helical filament after its MBP-tag is cleaved off and hence, I decided to set up the reported polymerisation assay to test if not only CARMA1 is able to increase polymerisation of BCL10, but also CARD9 to investigate what difference the additional coiled-coil fragments may have on the ability to nucleate the filaments.

The fluorescence polarisation assays performed in this study were based on the experiments performed by Qiao *et al.* (2013) [103]. Purification as well as labelling conditions were taken over in detail as described in the publication. In addition, I have been in contact with the group leader Professor Hao Wu to ensure purification of MBP-BCL10 and labelling were performed according to their protocols. However, as seen in section 5.5.4, the cleavage of the MBP-tag can be clearly followed, whereas polymerisation of BCL10 molecules cannot be observed within 2h nor after leaving the solution at room temperature overnight. Even the addition of CARMA1 to the FPA sample did not show any difference.

There are similarities between CARMA1/BCL10 and NOD2/RIP2 and similar to CARMA1, NOD2 oligomerises when activated. Furthermore, both their downstream effector domains are prone to aggregate heavily when being purified individually and soluble purification can only be achieved using a solubility enhancement tag. Therefore, the question arose if the new observation about BCL10 filament formation can be applied to RIP2-CARD, where NOD2 would nucleate and enhance RIP2 filament

formation? FPA experiments were designed and executed similar to BCL10/CARMA1 and are presented in section 4.4.3 of this work. However, very much to our surprise we could not observe RIP2-CARD polymerisation, although the MBP-tag had clearly been cleaved. Hence, similar as observed for CARMA1, the addition of NOD2 did not nucleate RIP2-CARD filament formation.

Our inability to reproduce these FPA experiments is supporting the critical voices suggesting that the recently reported CARD filaments may be an artefact of CARDS simply being overexpressed. In the literature there are contradicting reports of the subcellular distribution of overexpressed BCL10. Guet and Vito (2000) [105] have reported that BCL10 is localised in the cytoplasm as filamentous structures. On the other hand, Thome *et al.* (2001) [166] show only a granular staining pattern of BCL10 in the cytoplasm of transfected HeLa cells, whereas filamentous structures were only observed with a significant higher concentrations of the BCL10 expression vector. At present, we cannot explain why neither BCL10 nor RIP2 oligomerise in our assays, in contrast to published studies, and believe that further studies are required to fully understand the significance of fibre formation and its physiological role.

6.7 Future directions

In my thesis I wanted to address the question if there is a common mode of interaction in the CARD subfamily. We assumed, based on studies with other members of the death domain superfamily, that CARs may either form “canonical” 1:1 complexes or may form multi-protein assemblies as observed in the PIDDosome for example. However, during the course of this work another mechanism for assembly of death domain family member containing complexes was identified – the formation of filaments. Filament formation is still a rather controversial topic and a key question for future studies is to investigate how relevant filament formation is in a physiological context and identify how many other members of this domain superfamily may behave in such a fashion.

I also aimed to describe the intramolecular interaction between the tandem CARs of NOD2, and its intermolecular interaction with the CARD of RIP2 in molecular detail. While I was able to provide further insight into the binding surfaces involved in these interactions using mutagenesis combined with biochemical and biophysical assays, ultimately an atomic structure is required to fully understand their architecture. However, extensive crystallisation trials in our group have not been able to produce any crystals. These studies had focused on the human and mouse proteins and it might now be necessary to extend these studies to other proteins from organisms whose sequences might be more amenable to crystallisation. Alternatively, other methods such as using crystallisation chaperones may be required to reveal the structural secrets of the NOD2-RIP2 complex.

Bibliography

1. Delves PJ, Martin SJ, Burton DR, Roitt IM: **Essential Immunology**, 11th edn. Malden, Massachusetts, USA: Blackwell Publishing 2006.
2. Janeway CA, Jr., Medzhitov R: **Innate immune recognition**. *Annu Rev Immunol* 2002, **20**:197-216.
3. Medzhitov R: **Toll-like receptors and innate immunity**. *Nature Reviews Immunology* 2001, **1**(2):135-145.
4. Schulze-Luehrmann J, Ghosh S: **Antigen-receptor signaling to nuclear factor kappa B**. *Immunity* 2006, **25**(5):701-715.
5. Hashimoto C, Hudson KL, Anderson KV: **The *Toll* gene of drosophila, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane potein**. *Cell* 1988, **52**(2):269-279.
6. Nomura N, Miyajima N, Sazuka T, Tanaka A, Kawarabayasi Y, Sato S, Nagase T, Seki N, Ishikawa K-i, Tabata S: **Prediction of the Coding Sequences of Unidentified Human Genes. I. The Coding Sequences of 40 New Genes (KIAA0001-KIAA0040) Deduced by Analysis of Randomly Sampled cDNA Clones from Human Immature Myeloid Cell Line KG-1**. *DNA Research* 1994, **1**(1):27-35.
7. O'Neill LAJ: **The interleukin-1 receptor/Toll-like receptor superfamily: 10 years of progress**. *Immunological Reviews* 2008, **226**(1):10-18.
8. Yin Q, Fu T-M, Li J, Wu H: **Structural Biology of Innate Immunity**. *Annu Rev Immunol* 2015, **33**:393-416.
9. Kersse K, Bertrand MJ, Lamkanfi M, Vandenabeele P: **NOD-like receptors and the innate immune system: coping with danger, damage and death**. *Cytokine Growth Factor Rev* 2011, **22**(5-6):257-276.
10. Ruland J: **CARD9 Signaling in the Innate Immune Response**. *Annals of the New York Academy of Sciences* 2008, **1143**(1):35-44.
11. Akira S, Takeda K: **Toll-like receptor signalling**. *Nature Reviews Immunology* 2004, **4**(7):499-511.
12. Carpenter S, O'Neill Luke AJ: **Recent insights into the structure of Toll-like receptors and post-translational modifications of their associated signalling proteins**. *Biochemical Journal* 2009, **422**(1):1-10.
13. Burns K, Martinon F, Esslinger C, Pahl H, Schneider P, Bodmer J-L, Di Marco F, French L, Tschopp J: **MyD88, an Adapter Protein Involved in Interleukin-1 Signaling**. *Journal of Biological Chemistry* 1998, **273**(20):12203-12209.
14. Ohnishi H, Tochio H, Kato Z, Orii KE, Li A, Kimura T, Hiroaki H, Kondo N, Shirakawa M: **Structural basis for the multiple interactions of the MyD88 TIR domain in TLR4 signaling**. *Proceedings of the National Academy of Sciences* 2009, **106**(25):10260-10265.
15. Nyman T, Stenmark P, Flodin S, Johansson I, Hammarström M, Nordlund P: **The Crystal Structure of the Human Toll-like Receptor 10 Cytoplasmic Domain Reveals a Putative Signaling Dimer**. *Journal of Biological Chemistry* 2008, **283**(18):11861-11865.
16. Bell JK, Askins J, Hall PR, Davies DR, Segal DM: **The dsRNA binding site of human Toll-like receptor 3**. *Proceedings of the National Academy of Sciences* 2006, **103**(23):8792-8797.
17. Lin SC, Lo YC, Wu H: **Helical assembly in the MyD88-IRAK4-IRAK2 complex in TLR/IL-1R signalling**. *Nature* 2010, **465**(7300):885-890.
18. Berke IC, Li Y, Modis Y: **Structural basis of innate immune recognition of viral RNA**. *Cellular Microbiology* 2013, **15**(3):386-394.
19. K Drickamer a, Taylor ME: **Biology of Animal Lectins**. *Annual Review of Cell Biology* 1993, **9**(1):237-264.

20. Hardison SE, Brown GD: **C-type lectin receptors orchestrate antifungal immunity.** *Nat Immunol* 2012, **13**(9):817-822.
21. Zelensky AN, Gready JE: **The C-type lectin-like domain superfamily.** *Febs Journal* 2005, **272**(24):6179-6217.
22. Taylor PR, Tsoni SV, Willment JA, Dennehy KM, Rosas M, Findon H, Haynes K, Steele C, Botto M, Gordon S *et al*: **Dectin-1 is required for beta-glucan recognition and control of fungal infection.** *Nat Immunol* 2007, **8**(1):31-38.
23. Gross O, Gewies A, Finger K, Schäfer M, Sparwasser T, Peschel C, Förster I, Ruland J: **Card9 controls a non-TLR signalling pathway for innate anti-fungal immunity.** *Nature* 2006, **442**(7103):651-656.
24. Bertin J, Guo Y, Wang L, Srinivasula SM, Jacobson MD, Poyet JL, Merriam S, Du MQ, Dyer MJ, Robison KE *et al*: **CARD9 is a novel caspase recruitment domain-containing protein that interacts with BCL10/CLAP and activates NF-kappa B.** *The Journal of biological chemistry* 2000, **275**(52):41082-41086.
25. Lucas PC, Yonezumi M, Inohara N, McAllister-Lucas LM, Abazeed ME, Chen FF, Yamaoka S, Seto M, Núñez G: **Bcl10 and MALT1, Independent Targets of Chromosomal Translocation in MALT Lymphoma, Cooperate in a Novel NF-kB Signaling Pathway.** *Journal of Biological Chemistry* 2001, **276**(22):19012-19019.
26. Strasser D, Neumann K, Bergmann H, Marakalala Mohlopheni J, Guler R, Rojowska A, Hopfner K-P, Brombacher F, Urlaub H, Baier G *et al*: **Syk Kinase-Coupled C-type Lectin Receptors Engage Protein Kinase C-δ to Elicit Card9 Adaptor-Mediated Innate Immunity.** *Immunity* 2012, **36**(1):32-42.
27. Glocker E-O, Hennigs A, Nabavi M, Schäffer AA, Woellner C, Salzer U, Pfeifer D, Veelken H, Warnatz K, Tahami F *et al*: **A Homozygous CARD9 Mutation in a Family with Susceptibility to Fungal Infections.** *New England Journal of Medicine* 2009, **361**(18):1727-1735.
28. Meylan E, Tschopp J, Karin M: **Intracellular pattern recognition receptors in the host response.** *Nature* 2006, **442**(7098):39-44.
29. Saito T, Gale Jr M: **Principles of intracellular viral recognition.** *Current Opinion in Immunology* 2007, **19**(1):17-23.
30. Yoneyama M, Kikuchi M, Matsumoto K, Imaizumi T, Miyagishi M, Taira K, Foy E, Loo YM, Gale M, Akira S *et al*: **Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity.** *J Immunol* 2005, **175**(5):2851-2858.
31. Hornung V, Ellegast J, Kim S, Brzózka K, Jung A, Kato H, Poeck H, Akira S, Conzelmann K-K, Schlee M *et al*: **5'-Triphosphate RNA Is the Ligand for RIG-I.** *Science* 2006, **314**(5801):994-997.
32. Pichlmair A, Schulz O, Tan CP, Näslund TI, Liljeström P, Weber F, Reis e Sousa C: **RIG-I-Mediated Antiviral Responses to Single-Stranded RNA Bearing 5'-Phosphates.** *Science* 2006, **314**(5801):997-1001.
33. Kowalinski E, Lunardi T, McCarthy Andrew A, Luber J, Brunel J, Grigorov B, Gerlier D, Cusack S: **Structural Basis for the Activation of Innate Immune Pattern-Recognition Receptor RIG-I by Viral RNA.** *Cell* 2011, **147**(2):423-435.
34. Jiang F, Ramanathan A, Miller MT, Tang G-Q, Gale M, Patel SS, Marcotrigiano J: **Structural basis of RNA recognition and activation by innate immune receptor RIG-I.** *Nature* 2011, **479**(7373):423-427.
35. Luo D, Ding Steve C, Vela A, Kohlway A, Lindenbach Brett D, Pyle Anna M: **Structural Insights into RNA Recognition by RIG-I.** *Cell* 2011, **147**(2):409-422.
36. Gack MU, Shin YC, Joo C-H, Urano T, Liang C, Sun L, Takeuchi O, Akira S, Chen Z, Inoue S *et al*: **TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity.** *Nature* 2007, **446**(7138):916-920.
37. Peisley A, Wu B, Xu H, Chen ZJ, Hur S: **Structural basis for ubiquitin-mediated antiviral signal activation by RIG-I.** *Nature* 2014, **509**(7498):110-114.

38. Jiang X, Kinch LN, Brautigam Chad A, Chen X, Du F, Grishin NV, Chen Zhijian J: **Ubiquitin-Induced Oligomerization of the RNA Sensors RIG-I and MDA5 Activates Antiviral Innate Immune Response.** *Immunity* 2012, **36**(6):959-973.
39. Xu H, He X, Zheng H, Huang LJ, Hou F, Yu Z, de la Cruz MJ, Borkowski B, Zhang X, Chen ZJ *et al*: **Structural basis for the prion-like MAVS filaments in antiviral innate immunity**, vol. 3; 2014.
40. Hou F, Sun L, Zheng H, Skaug B, Jiang Q-X, Chen Zhijian J: **MAVS Forms Functional Prion-like Aggregates to Activate and Propagate Antiviral Innate Immune Response.** *Cell* 2011, **146**(3):448-461.
41. Kato H, Takeuchi O, Mikamo-Satoh E, Hirai R, Kawai T, Matsushita K, Hiiragi A, Dermody TS, Fujita T, Akira S: **Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid-inducible gene-I and melanoma differentiation-associated gene 5.** *J Exp Med* 2008, **205**(7):1601-1610.
42. Berke IC, Modis Y: **MDA5 cooperatively forms dimers and ATP-sensitive filaments upon binding double-stranded RNA.** *The EMBO Journal* 2012, **31**(7):1714-1726.
43. Berke IC, Yu X, Modis Y, Egelman EH: **MDA5 assembles into a polar helical filament on dsRNA.** *P Natl Acad Sci USA* 2012, **109**(45):18437-18441.
44. Wu B, Peisley A, Richards C, Yao H, Zeng X, Lin C, Chu F, Walz T, Hur S: **Structural Basis for dsRNA Recognition, Filament Formation, and Antiviral Signal Activation by MDA5.** *Cell* 2013, **152**(1-2):276-289.
45. Peisley A, Wu B, Yao H, Walz T, Hur S: **RIG-I Forms Signaling-Competent Filaments in an ATP-Dependent, Ubiquitin-Independent Manner.** *Mol Cell* 2013, **51**(5):573-583.
46. Ting J, Lovering R, Alnemri E, Bertin J, Boss J, Davis B, Flavell R, Girardin S, Godzik A, Harton J: **The NLR gene family: a standard nomenclature.** *Immunity* 2008, **28**:285 - 287.
47. Strowig T, Henao-Mejia J, Elinav E, Flavell R: **Inflammasomes in health and disease.** *Nature* 2012, **481**(7381):278-286.
48. Elinav E, Strowig T, Kau Andrew L, Henao-Mejia J, Thaiss Christoph A, Booth Carmen J, Peaper David R, Bertin J, Eisenbarth Stephanie C, Gordon Jeffrey I *et al*: **NLRP6 Inflammasome Regulates Colonic Microbial Ecology and Risk for Colitis.** *Cell* 2011, **145**(5):745-757.
49. Franchi L, Eigenbrod T, Munoz-Planillo R, Nunez G: **The inflammasome: a caspase-1-activation platform that regulates immune responses and disease pathogenesis.** *Nat Immunol* 2009, **10**(3):241-247.
50. Martinon F, Tschopp J: **Inflammatory Caspases: Linking an Intracellular Innate Immune System to Autoinflammatory Diseases.** *Cell* 2004, **117**(5):561-574.
51. Rathinam VAK, Vanaja SK, Fitzgerald KA: **Regulation of inflammasome signaling.** *Nat Immunol* 2012, **13**(4):333-332.
52. Neer EJ, Schmidt CJ, Nambudripad R, Smith TF: **The ancient regulatory-protein family of WD-repeat proteins.** *Nature* 1994, **371**(6495):297-300.
53. Reubold Thomas F, Wohlgemuth S, Eschenburg S: **Crystal Structure of Full-Length Apaf-1: How the Death Signal Is Relayed in the Mitochondrial Pathway of Apoptosis.** *Structure* 2011, **19**(8):1074-1083.
54. Yuan S, Yu X, Topf M, Ludtke SJ, Wang X, Akey CW: **Structure of an Apoptosome-Procaspase-9 CARD Complex.** *Structure* 2010, **18**(5):571-583.
55. Yuan S, Akey Christopher W: **Apoptosome Structure, Assembly, and Procaspase Activation.** *Structure* 2013, **21**(4):501-515.
56. Faustin B, Lartigue L, Bruey J, Luciano F, Sergienko E, Bailly-Maitre B, Volkmann N, Hanein D, Rouiller I, Reed J: **Reconstituted NALP1 inflammasome reveals two-step mechanism of caspase 1 activation.** *Mol Cell* 2007, **25**:713 - 724.
57. Shenoy AR, Wellington DA, Kumar P, Kassa H, Booth CJ, Cresswell P, MacMicking JD: **GBP5 Promotes NLRP3 Inflammasome Assembly and Immunity in Mammals.** *Science* 2012, **336**(6080):481-485.

58. Rathinam Vijay AK, Vanaja Sivapriya K, Waggoner L, Sokolovska A, Becker C, Stuart Lynda M, Leong John M, Fitzgerald Katherine A: **TRIF Licenses Caspase-11-Dependent NLRP3 Inflammasome Activation by Gram-Negative Bacteria.** *Cell* 2012, **150**(3):606-619.
59. Khare S, Dorfleutner A, Bryan Nicole B, Yun C, Radian Alexander D, de Almeida L, Rojanasakul Y, Stehlik C: **An NLRP7-Containing Inflammasome Mediates Recognition of Microbial Lipopeptides in Human Macrophages.** *Immunity* 2012, **36**(3):464-476.
60. Vladimer Gregory I, Weng D, Paquette Sara WM, Vanaja Sivapriya K, Rathinam Vijay AK, Aune Marie H, Conlon Joseph E, Burbage Joseph J, Proulx Megan K, Liu Q *et al*: **The NLRP12 Inflammasome Recognizes Yersinia pestis.** *Immunity* 2012, **37**(1):96-107.
61. Qu Y, Misaghi S, Izrael-Tomasevic A, Newton K, Gilmour LL, Lamkanfi M, Louie S, Kayagaki N, Liu J, Komuves L *et al*: **Phosphorylation of NLRC4 is critical for inflammasome activation.** *Nature* 2012, **490**(7421):539-542.
62. Lu A, Magupalli Venkat G, Ruan J, Yin Q, Atianand Maninjay K, Vos MR, Schröder Gunnar F, Fitzgerald Katherine A, Wu H, Egelman Edward H: **Unified Polymerization Mechanism for the Assembly of ASC-Dependent Inflammasomes.** *Cell* 2014, **156**(6):1193-1206.
63. Philpott DJ, Sorbara MT, Robertson SJ, Croitoru K, Girardin SE: **NOD proteins: regulators of inflammation in health and disease.** *Nat Rev Immunol* 2014, **14**(1):9-23.
64. Ogura Y, Inohara N, Benito A, Chen FF, Yamaoka S, Nunez G: **Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-kappa B.** *Journal of Biological Chemistry* 2001, **276**(7):4812-4818.
65. Inohara N, Ogura Y, Fontalba A, Gutierrez O, Pons F, Crespo J, Fukase K, Inamura S, Kusumoto S, Hashimoto M *et al*: **Host recognition of bacterial muramyl dipeptide mediated through NOD2.** *Journal of Biological Chemistry* 2003, **278**(8):5509-5512.
66. Chamaillard M, Hashimoto M, Horie Y, Masumoto J, Qiu S, Saab L, Ogura Y, Kawasaki A, Fukase K, Kusumoto S *et al*: **An essential role for NOD1 in host recognition of bacterial peptidoglycan containing diaminopimelic acid.** *Nat Immunol* 2003, **4**(7):702-707.
67. Mo J, Boyle JP, Howard CB, Monie TP, Davis BK, Duncan JA: **Pathogen Sensing by Nucleotide-binding Oligomerization Domain-containing Protein 2 (NOD2) Is Mediated by Direct Binding to Muramyl Dipeptide and ATP.** *Journal of Biological Chemistry* 2012, **287**(27):23057-23067.
68. Grimes CL, Ariyananda LDZ, Melnyk JE, O'Shea EK: **The Innate Immune Protein Nod2 Binds Directly to MDP, a Bacterial Cell Wall Fragment.** *Journal of the American Chemical Society* 2012, **134**(33):13535-13537.
69. Strober W, Murray PJ, Kitani A, Watanabe T: **Signalling pathways and molecular interactions of NOD1 and NOD2.** *Nat Rev Immunol* 2006, **6**(1):9-20.
70. Ubersax JA, Ferrell Jr JE: **Mechanisms of specificity in protein phosphorylation.** *Nat Rev Mol Cell Biol* 2007, **8**(7):530-541.
71. Canning P, Ruan Q, Schwerd T, Hrdinka M, Maki Jenny L, Saleh D, Suebsuwong C, Ray S, Brennan Paul E, Cuny Gregory D *et al*: **Inflammatory Signaling by NOD-RIPK2 Is Inhibited by Clinically Relevant Type II Kinase Inhibitors.** *Chemistry & Biology* 2015, **22**(9):1174-1184.
72. Inohara N, Koseki T, del Peso L, Hu Y, Yee C, Chen S, Carrio R, Merino J, Liu D, Ni J *et al*: **Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor-kappaB.** *The Journal of biological chemistry* 1999, **274**(21):14560-14567.
73. Windheim M, Lang C, Pegg M, Plater Lorna A, Cohen P: **Molecular mechanisms involved in the regulation of cytokine production by muramyl dipeptide.** *Biochemical Journal* 2007, **404**(2):179-190.

74. Park J-H, Kim Y-G, McDonald C, Kanneganti T-D, Hasegawa M, Body-Malapel M, Inohara N, Núñez G: **RICK/RIP2 Mediates Innate Immune Responses Induced through Nod1 and Nod2 but Not TLRs.** *The Journal of Immunology* 2007, **178**(4):2380-2386.
75. Yang S, Wang B, Humphries F, Jackson R, Healy ME, Bergin R, Aviello G, Hall B, McNamara D, Darby T *et al*: **Pellino3 ubiquitinates RIP2 and mediates Nod2-induced signaling and protective effects in colitis.** *Nat Immunol* 2013, **14**(9):927-936.
76. Hasegawa M, Fujimoto Y, Lucas PC, Nakano H, Fukase K, Núñez G, Inohara N: **A critical role of RICK/RIP2 polyubiquitination in Nod-induced NF- κ B activation.** *The EMBO Journal* 2008, **27**(2):373-383.
77. Girardin SE, Tournebise R, Mavris M, Page AL, Li X, Stark GR, Bertin J, DiStefano PS, Yaniv M, Sansonetti PJ *et al*: **CARD4/Nod1 mediates NF- κ B and JNK activation by invasive *Shigella flexneri*.** *EMBO Rep* 2001, **2**(8):736-742.
78. Hsu Y-MS, Zhang Y, You Y, Wang D, Li H, Duramad O, Qin X-F, Dong C, Lin X: **The adaptor protein CARD9 is required for innate immune responses to intracellular pathogens.** *Nat Immunol* 2007, **8**(2):198-205.
79. Zhernakova A, Festen EM, Franke L, Trynka G, van Diemen CC, Monsuur AJ, Bevova M, Nijmeijer RM, van 't Slot R, Heijmans R *et al*: **Genetic Analysis of Innate Immunity in Crohn's Disease and Ulcerative Colitis Identifies Two Susceptibility Loci Harboring CARD9 and IL18RAP.** *The American Journal of Human Genetics* 2008, **82**(5):1202-1210.
80. Parkhouse R, Boyle JP, Mayle S, Sawmynaden K, Rittinger K, Monie TP: **Interaction between NOD2 and CARD9 involves the NOD2 NACHT and the linker region between the NOD2 CARDs and NACHT domain.** *Febs Letters* 2014, **588**(17):2830-2836.
81. Roth S, Ruland J: **Caspase recruitment domain-containing protein 9 signaling in innate immunity and inflammation.** *Trends Immunol* 2013, **34**(6):243-250.
82. Hayden MS, Ghosh S: **Shared Principles in NF- κ B Signaling.** *Cell* 2008, **132**(3):344-362.
83. Hacker H, Karin M: **Regulation and function of IKK and IKK-related kinases.** *Sci STKE* 2006, **2006**(357):re13-re13.
84. Hayden MS, Ghosh S: **NF- κ B, the first quarter-century: remarkable progress and outstanding questions.** *Genes & Development* 2012, **26**(3):203-234.
85. Tamura T, Yanai H, Savitsky D, Taniguchi T: **The IRF Family Transcription Factors in Immunity and Oncogenesis.** *Annu Rev Immunol* 2008, **26**(1):535-584.
86. Eguchi J, Yan Q-W, Schones DE, Kamal M, Hsu C-H, Zhang MQ, Crawford GE, Rosen ED: **Interferon Regulatory Factors Are Transcriptional Regulators of Adipogenesis.** *Cell Metabolism* 2008, **7**(1):86-94.
87. Sharma S, tenOever BR, Grandvaux N, Zhou G-P, Lin R, Hiscott J: **Triggering the Interferon Antiviral Response Through an IKK-Related Pathway.** *Science* 2003, **300**(5622):1148-1151.
88. Sato M, Suemori H, Hata N, Asagiri M, Ogasawara K, Nakao K, Nakaya T, Katsuki M, Noguchi S, Tanaka N *et al*: **Distinct and Essential Roles of Transcription Factors IRF-3 and IRF-7 in Response to Viruses for IFN- α/β Gene Induction.** *Immunity* 2000, **13**(4):539-548.
89. Treanor B: **B-cell receptor: from resting state to activate.** *Immunology* 2012, **136**(1):21-27.
90. Reth M: **Antigen receptor tail clue.** *Nature* 1989, **338**(6214):383-384.
91. Davis MM: **A New Trigger for T Cells.** *Cell* 2002, **110**(3):285-287.
92. Borst J, Coligan JE, Oettgen H, Pessano S, Malin R, Terhorst C: **The delta-chain and epsilon-chain of the human T3/T-cell receptor complex are distinct polypeptides.** *Nature* 1984, **312**(5993):455-458.
93. Samelson LE, Patel MD, Weissman AM, Harford JB, Klausner RD: **Antigen activation of murine T cells induces tyrosine phosphorylation of a polypeptide associated with the T cell antigen receptor.** *Cell* 1986, **46**(7):1083-1090.
94. Love PE, Hayes SM: **ITAM-mediated Signaling by the T-Cell Antigen Receptor.** *Csh Perspect Biol* 2010, **2**(6).

95. Geahlen RL: **Syk and pTyr'd: Signaling through the B cell antigen receptor.** *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 2009, **1793**(7):1115-1127.
96. Thome M, Charton JE, Pelzer C, Hailfinger S: **Antigen receptor signaling to NF-kappaB via CARMA1, BCL10, and MALT1.** *Csh Perspect Biol* 2010, **2**(9):a003004.
97. Bertin J, Wang L, Guo Y, Jacobson MD, Poyet JL, Srinivasula SM, Merriam S, DiStefano PS, Alnemri ES: **CARD11 and CARD14 are novel caspase recruitment domain (CARD)/membrane-associated guanylate kinase (MAGUK) family members that interact with BCL10 and activate NF-kappa B.** *The Journal of biological chemistry* 2001, **276**(15):11877-11882.
98. Pomerantz JL, Denny EM, Baltimore D: **CARD11 mediates factor-specific activation of NF-kB by the T cell receptor complex.** *The EMBO Journal* 2002, **21**(19):5184-5194.
99. Dimitratos SD, Woods DF, Stathakis DG, Bryant PJ: **Signaling pathways are focused at specialized regions of the plasma membrane by scaffolding proteins of the MAGUK family.** *BioEssays* 1999, **21**(11):912-921.
100. Hara H, Yokosuka T, Hirakawa H, Ishihara C, Yasukawa S, Yamazaki M, Koseki H, Yoshida H, Saito T: **Clustering of CARMA1 through SH3-GUK domain interactions is required for its activation of NF-kB signalling.** *Nat Commun* 2015, **6**.
101. Matsumoto R, Wang D, Blonska M, Li H, Kobayashi M, Pappu B, Chen Y, Lin X: **Phosphorylation of CARMA1 plays a critical role in T Cell receptor-mediated NF-kappaB activation.** *Immunity* 2005, **23**(6):575-585.
102. Sommer K, Guo B, Pomerantz JL, Bandaranayake AD, Moreno-Garcia ME, Ovechkina YL, Rawlings DJ: **Phosphorylation of the CARMA1 linker controls NF-kappaB activation.** *Immunity* 2005, **23**(6):561-574.
103. Qiao Q, Yang C, Zheng C, Fontán L, David L, Yu X, Bracken C, Rosen M, Melnick A, Egelman Edward H *et al*: **Structural Architecture of the CARMA1/Bcl10/MALT1 Signalingosome: Nucleation-Induced Filamentous Assembly.** *Mol Cell* 2013, **51**(6):766-779.
104. Che TJ, You Y, Wang DH, Tanner MJ, Dixit VM, Lin X: **MALT1/paracaspase is a signaling component downstream of CARMA1 and mediates T cell receptor-induced NF-kappa B activation.** *Journal of Biological Chemistry* 2004, **279**(16):15870-15876.
105. Guet C, Vito P: **Caspase Recruitment Domain (Card-Dependent) Cytoplasmic Filaments Mediate Bcl10-Induced Nf-kb Activation.** *The Journal of Cell Biology* 2000, **148**(6):1131-1140.
106. Rossman JS, Stoicheva NG, Langel FD, Patterson GH, Lippincott-Schwartz J, Schaefer BC: **POLKADOTS Are Foci of Functional Interactions in T-Cell Receptor-mediated Signaling to NF-kB.** *Molecular Biology of the Cell* 2006, **17**(5):2166-2176.
107. Oeckinghaus A, Wegener E, Welteke V, Ferch U, Arslan SÇ, Ruland J, Scheidereit C, Krappmann D: **Malt1 ubiquitination triggers NF-kB signaling upon T-cell activation.** *The EMBO Journal* 2007, **26**(22):4634-4645.
108. Thome M: **Multifunctional roles for MALT1 in T-cell activation.** *Nat Rev Immunol* 2008, **8**(7):495-500.
109. Hu S, Vincenz C, Buller M, Dixit VM: **A Novel Family of Viral Death Effector Domain-containing Molecules That Inhibit Both CD-95- and Tumor Necrosis Factor Receptor-1-induced Apoptosis.** *Journal of Biological Chemistry* 1997, **272**(15):9621-9624.
110. Kersse K, Verspurten J, Vanden Berghe T, Vandenabeele P: **The death-fold superfamily of homotypic interaction motifs.** *Trends in biochemical sciences* 2011, **36**(10):541-552.
111. Park HH, Lo YC, Lin SC, Wang L, Yang JK, Wu H: **The death domain superfamily in intracellular signaling of apoptosis and inflammation.** *Annu Rev Immunol* 2007, **25**:561-586.
112. Chou JJ, Matsuo H, Duan H, Wagner G: **Solution structure of the RAIDD CARD and model for CARD/CARD interaction in caspase-2 and caspase-9 recruitment.** *Cell* 1998, **94**(2):171-180.

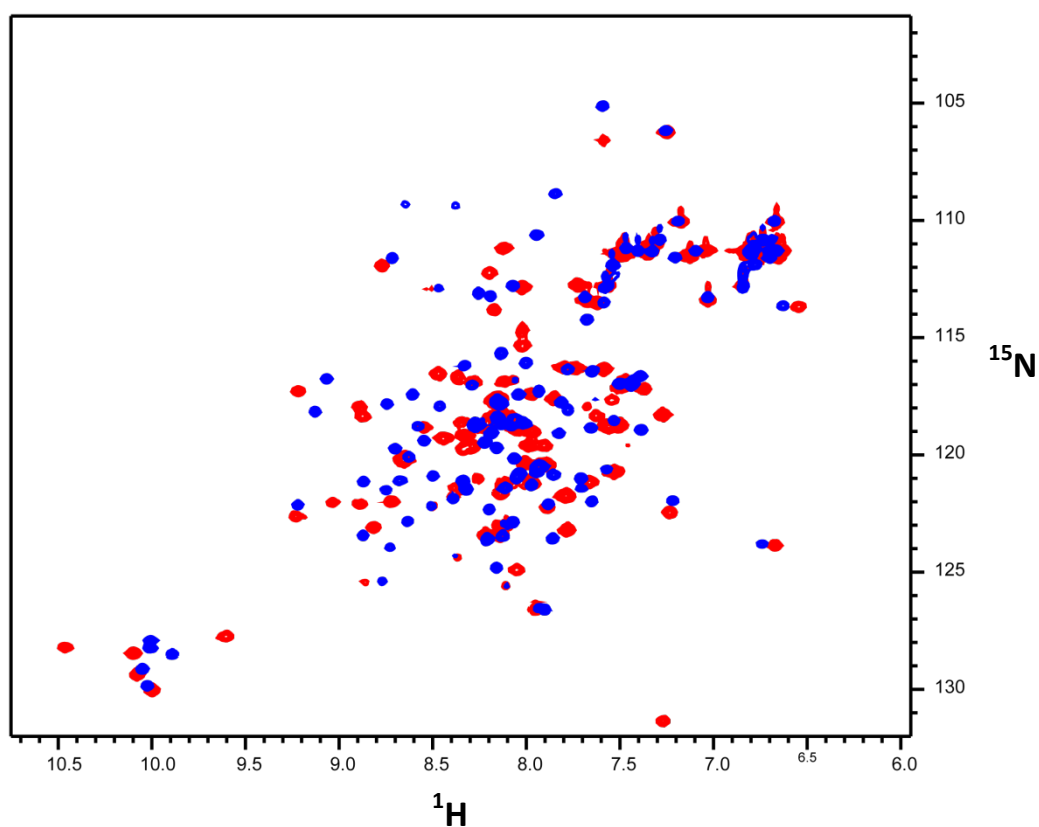
113. Eberstadt M, Huang BH, Chen ZH, Meadows RP, Ng SC, Zheng LX, Lenardo MJ, Fesik SW: **NMR structure and mutagenesis of the FADD (Mort1) death-effector domain.** *Nature* 1998, **392**(6679):941-945.
114. Hiller S, Kohl A, Fiorito F, Herrmann T, Wider G, Tschopp J, Grutter MG, Wuthrich K: **NMR structure of the apoptosis- and inflammation-related NALP1 pyrin domain.** *Structure* 2003, **11**(10):1199-1205.
115. Huang BH, Eberstadt M, Olejniczak ET, Meadows RP, Fesik SW: **NMR structure and mutagenesis of the Fas (APO-1/CD95) death domain.** *Nature* 1996, **384**(6610):638-641.
116. Qin HX, Srinivasula SM, Wu G, Fernandes-Alnemri T, Alnemri ES, Shi YG: **Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1.** *Nature* 1999, **399**(6736):549-557.
117. Weber CH, Vincenz C: **The death domain superfamily: a tale of two interfaces?** *Trends in Biochemical Sciences* 2001, **26**(8):475-481.
118. Jeong E-J, Bang S, Lee TH, Park YI, Sim W-S, Kim K-S: **The Solution Structure of FADD Death Domain: STRUCTURAL BASIS OF DEATH DOMAIN INTERACTIONS OF Fas AND FADD.** *Journal of Biological Chemistry* 1999, **274**(23):16337-16342.
119. Sukits SF, Lin L-L, Hsu S, Malakian K, Powers R, Xu G-Y: **Solution structure of the tumor necrosis factor receptor-1 death domain1.** *J Mol Biol* 2001, **310**(4):895-906.
120. Janssens S, Tinel A: **The PIDDosome, DNA-damage-induced apoptosis and beyond.** *Cell Death Differ* 2012, **19**(1):13-20.
121. Yang JK, Wang L, Zheng L, Wan F, Ahmed M, Lenardo MJ, Wu H: **Crystal Structure of MC159 Reveals Molecular Mechanism of DISC Assembly and FLIP Inhibition.** *Mol Cell* 2005, **20**(6):939-949.
122. Hill JM, Vaidyanathan H, Ramos JW, Ginsberg MH, Werner MH: **Recognition of ERK MAP kinase by PEA-15 reveals a common docking site within the death domain and death effector domain.** *The EMBO Journal* 2002, **21**(23):6494-6504.
123. Shen C, Yue H, Pei J, Guo X, Wang T, Quan J-M: **Crystal structure of the death effector domains of caspase-8.** *Biochemical and Biophysical Research Communications* 2015, **463**(3):297-302.
124. Park HH, Lo YC, Lin SC, Wang L, Yang JK, Wu H: **The death domain superfamily in intracellular signaling of apoptosis and inflammation.** In: *Annu Rev Immunol.* vol. 25. Palo Alto: Annual Reviews; 2007: 561-586.
125. Eliezer D: **Folding Pyrin into the Family.** *Structure* 2003, **11**(10):1190-1191.
126. de Alba E: **Structure and interdomain dynamics of apoptosis-associated speck-like protein containing a CARD (ASC).** *J Biol Chem* 2009, **284**(47):32932-32941.
127. Kao W-P, Yang C-Y, Su T-W, Wang Y-T, Lo Y-C, Lin S-C: **The versatile roles of CARDs in regulating apoptosis, inflammation, and NF- κ B signaling.** *Apoptosis* 2015, **20**(2):174-195.
128. Yoneyama M, Kikuchi M, Natsukawa T, Shinobu N, Imaizumi T, Miyagishi M, Taira K, Akira S, Fujita T: **The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses.** *Nat Immunol* 2004, **5**(7):730-737.
129. Kovacsics M, Martinon F, Micheau O, Bodmer J-L, Hofmann K, Tschopp J: **Overexpression of Helicard, a CARD-Containing Helicase Cleaved during Apoptosis, Accelerates DNA Degradation.** *Current Biology* 2002, **12**(10):838-843.
130. Wagner RN, Proell M, Kufer TA, Schwarzenbacher R: **Evaluation of Nod-Like Receptor (NLR) Effector Domain Interactions.** *Plos One* 2009, **4**(4):e4931.
131. Fridh V, Rittinger K: **The Tandem CARDs of NOD2: Intramolecular Interactions and Recognition of RIP2.** *Plos One* 2012, **7**(3):e34375.
132. Meyer B, Peters T: **NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors.** *Angewandte Chemie International Edition* 2003, **42**(8):864-890.

133. Wu B, Peisley A, Tetrault D, Li Z, Egelman Edward H, Magor Katharine E, Walz T, Penczek Pawel A, Hur S: **Molecular Imprinting as a Signal-Activation Mechanism of the Viral RNA Sensor RIG-I.** *Mol Cell* 2014, **55**(4):511-523.
134. Aslanidis C, Dejong PJ: **Ligation-independent cloning of PCR products (LIC-PCR).** *Nucleic Acids Research* 1990, **18**(20):6069-6074.
135. Mach H, Middaugh CR, Denslow N: **Determining the Identity and Purity of Recombinant Proteins by UV Absorption Spectroscopy.** In: *Current Protocols in Protein Science*. John Wiley & Sons, Inc.; 2001.
136. Pierce MM, Raman CS, Nall BT: **Isothermal Titration Calorimetry of Protein-Protein Interactions.** *Methods* 1999, **19**(2):213-221.
137. Velázquez Campoy A, Freire E: **ITC in the post-genomic era...? Priceless.** *Biophysical Chemistry* 2005, **115**(2-3):115-124.
138. Doyle ML: **Titration Microcalorimetry.** In: *Current Protocols in Protein Science*. John Wiley & Sons, Inc.; 2001.
139. Prabhu NV, Sharp KA: **HEAT CAPACITY IN PROTEINS.** *Annual Review of Physical Chemistry* 2005, **56**(1):521-548.
140. Wiseman T, Williston S, Brandts JF, Lin L-N: **Rapid measurement of binding constants and heats of binding using a new titration calorimeter.** *Analytical Biochemistry* 1989, **179**(1):131-137.
141. Martin SR, Schilstra MJ: **Circular Dichroism and Its Application to the Study of Biomolecules.** In: *Methods in Cell Biology*. vol. Volume 84: Academic Press; 2008: 263-293.
142. Ericsson UB, Hallberg BM, DeTitta GT, Dekker N, Nordlund P: **Thermofluor-based high-throughput stability optimization of proteins for structural studies.** *Analytical Biochemistry* 2006, **357**(2):289-298.
143. Wen J, Arakawa T, Philo JS: **Size-Exclusion Chromatography with On-Line Light-Scattering, Absorbance, and Refractive Index Detectors for Studying Proteins and Their Interactions.** *Analytical Biochemistry* 1996, **240**(2):155-166.
144. Lakowicz JR: **Principles of Fluorescence Spectroscopy**, 2nd edn. New York: Springer US; 2006.
145. Cole C, Barber JD, Barton GJ: **The Jpred 3 secondary structure prediction server.** *Nucleic Acids Research* 2008, **36**:W197-W201.
146. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J *et al*: **Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.** *Molecular Systems Biology* 2011, **7**(1).
147. Smith DB, Johnson KS: **Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase.** *Gene* 1988, **67**(1):31-40.
148. Hasegawa H, Holm L: **Advances and pitfalls of protein structural alignment.** *Current Opinion in Structural Biology* 2009, **19**(3):341-348.
149. Schrodinger, LLC: **The PyMOL Molecular Graphics System, Version 1.3r1.** In.; 2010.
150. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ: **Jalview Version 2—a multiple sequence alignment editor and analysis workbench.** *Bioinformatics* 2009, **25**(9):1189-1191.
151. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Research* 2004, **32**(5):1792-1797.
152. Arnold K, Bordoli L, Kopp J, Schwede T: **The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.** *Bioinformatics* 2006, **22**(2):195-201.
153. Otting G: **Experimental NMR techniques for studies of protein-ligand interactions.** *Current Opinion in Structural Biology* 1993, **3**(5):760-768.
154. Levine LM, Michener ML, Toth MV, Holwerda BC: **Measurement of Specific Protease Activity Utilizing Fluorescence Polarization.** *Analytical Biochemistry* 1997, **247**(1):83-88.

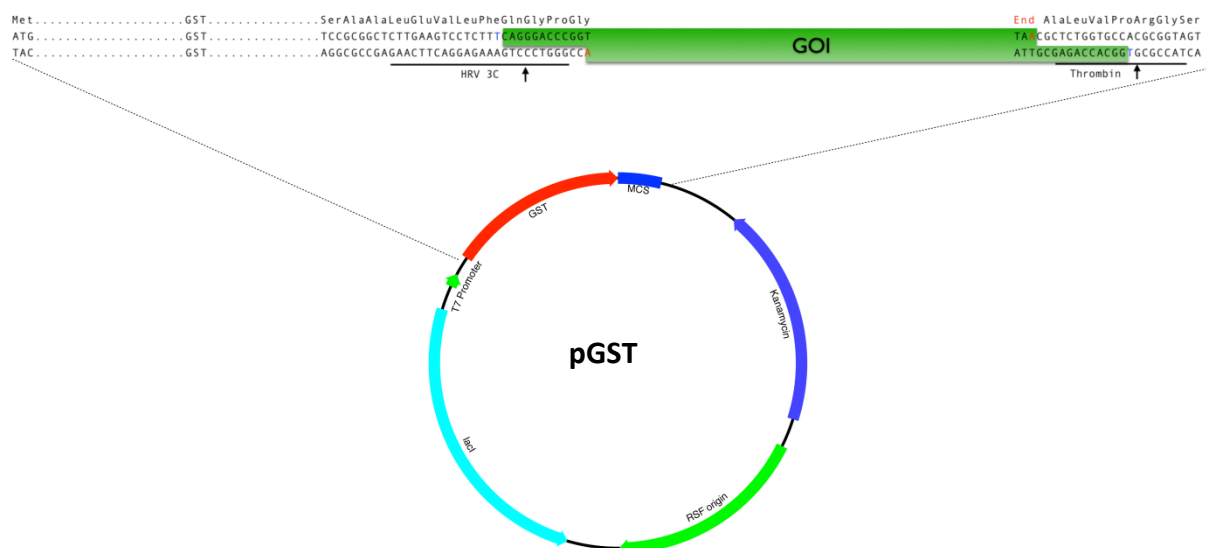
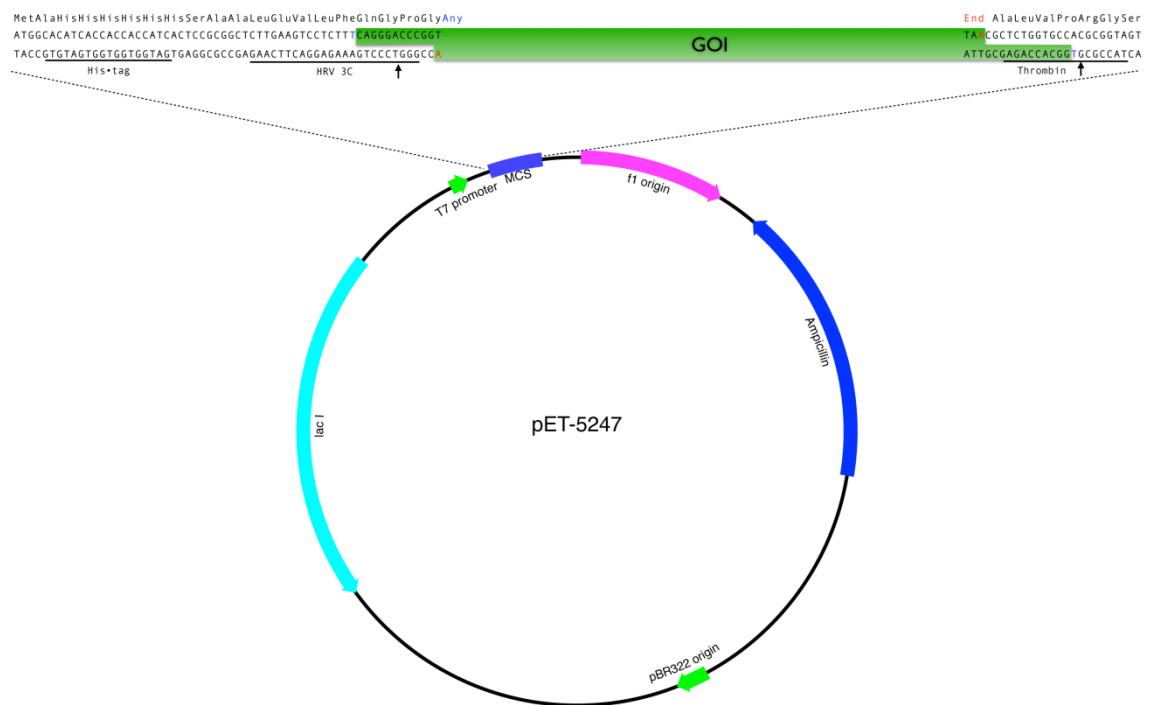
155. Sattler M, Schleucher J, Griesinger C: **Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients.** *Progress in Nuclear Magnetic Resonance Spectroscopy* 1999, **34**(2):93-158.
156. Li S, Yang X, Shao J, Shen Y: **Structural Insights into the Assembly of CARMA1 and BCL10.** *Plos One* 2012, **7**(8).
157. Wolf E, Kim PS, Berger B: **MultiCoil: A program for predicting two-and three-stranded coiled coils.** *Protein Science* 1997, **6**(6):1179-1189.
158. Mooij WTM, Mitsiki E, Perrakis A: **ProteinCCD: enabling the design of protein truncation constructs for expression and crystallization experiments.** *Nucleic Acids Research* 2009.
159. Lupas A: **Predicting coiled-coil regions in proteins.** *Current Opinion in Structural Biology* 1997, **7**(3):388-393.
160. Jang T-h, Park HH: **Generalized semi-refolding methods for purification of the functional death domain superfamily.** *Journal of Biotechnology* 2011, **151**(4):335-342.
161. WANG C, CASTRO AF, WILKES DM, ALTENBERG GA: **Expression and purification of the first nucleotide-binding domain and linker region of human multidrug resistance gene product: comparison of fusions to glutathione S-transferase, thioredoxin and maltose-binding protein.** *Biochemical Journal* 1999, **338**(1):77-81.
162. Boyle J, Mayle S, Parkhouse R, Monie TP: **Comparative genomic and sequence analysis provides insight into the molecular functionality of NOD1 and NOD2.** *Frontiers in Immunology* 2013, **4**.
163. Manon F, Favier A, Nunez G, Simorre JP, Cusack S: **Solution structure of NOD1 CARD and mutational analysis of its interaction with the CARD of downstream kinase RICK.** *J Mol Biol* 2007, **365**(1):160-174.
164. Tanabe T, Chamailard M, Ogura Y, Zhu L, Qiu S, Masumoto J, Ghosh P, Moran A, Predergast MM, Tromp G *et al*: **Regulatory regions and critical residues of NOD2 involved in muramyl dipeptide recognition.** *The EMBO Journal* 2004, **23**(7):1587-1597.
165. Rosenstiel P, Huse K, Till A, Hampe J, Hellmig S, Sina C, Billmann S, von Kampen O, Waetzig G, Platzer M: **A short isoform of NOD2/CARD15, NOD2-S, is an endogenous inhibitor of NOD2/receptor-interacting protein kinase 2-induced signaling pathways.** *Proc Natl Acad Sci USA* 2006, **103**:3280 - 3285.
166. Thome M, Gaide O, Micheau O, Martinon F, Bonnet D, Gonzalez M, Tschopp J: **Equine Herpesvirus Protein E10 Induces Membrane Recruitment and Phosphorylation of Its Cellular Homologue, Bcl-10.** *The Journal of Cell Biology* 2001, **152**(5):1115-1122.
167. <http://www.urbandictionary.com/define.php?term=phd>

Appendix

A 1-1 NMR titration experiment of NOD2 tandem CARDs. ^1H - ^{15}N -TROSY-HSQC of ^{15}N -CARDa alone (blue) and in complex with NOD2-CARDb (red). Spectra recorded at 25 °C. Buffer: 100 mM NaCl, 20 mM Na-Phosphate pH 7.1 and 2 mM TCEP.

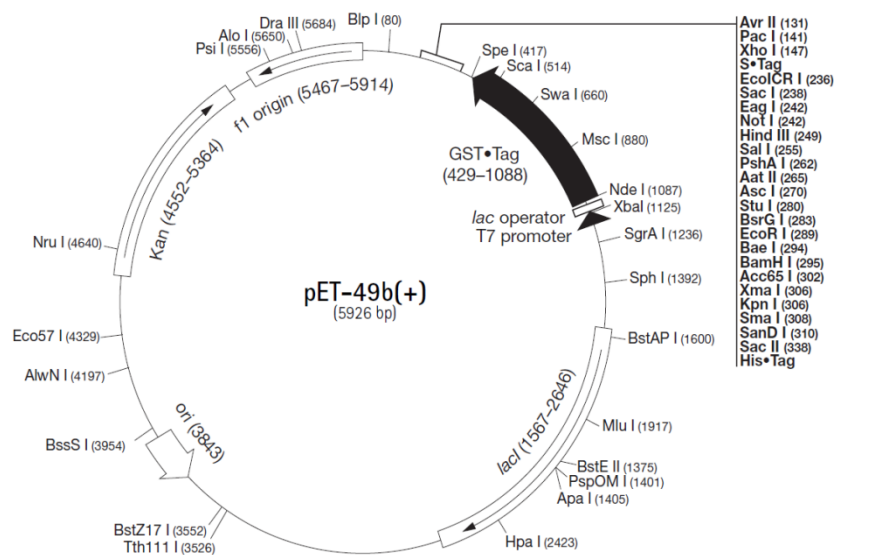
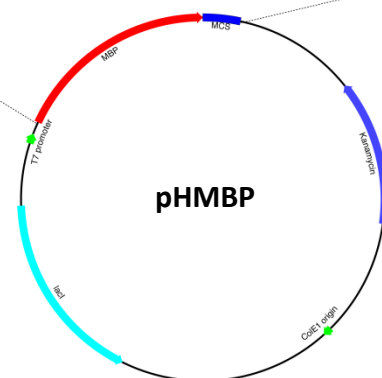


A 2-1 Expression vectors used in this study. Figure for pET52-47, pGST and pHMBP were kindly provided by Vangelis Christodoulou. Figure showing pET49b was adapted from <http://www.merckmillipore.com>.



Met.....MBP.....SerAlaAlaLeuGluValLeuPheGlnGlyProGly
 ATG.....MBP.....TCGCGGGCTCTGAAGTCCTCTTTCAGGGACCCGGT
 TAC.....MBP.....AGGCGCCGAGAACTTCAGGAGAAAGTCCTGGGCCC
 End AlaLeuValProArgGlySer
 TATCGCTCTGGTGCCACGGGGTAGT
 ATTCGAGACCAAGTGCGCATCA

HRV 3C ↑ GOI Thrombin ↑



T7 Promoter Primer
 #69348-3
 TCTCGATCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATCCCTCTAGAAATAATTTGTTAACTTAAGAAGGAGAT

Nde I
 ATACATATGTCCTTACTAGGTTATTGGAAAATTAAG...606 bp...GACCATCTCCAAATCGGATGGTTCAACTAGTGGTGGCGGTTCT
 MetSerProIleLeuGlyTyrTrpLysIleLys...202 aa...AspHisProProLysSerAspGlySerThrSerGlyGlyGlyGlySer

Acc65 I
 Kpn I
 Bae I
 SanD I
 Xma I
 Sma I
 His*Tag
 Sac II
 AATAACAATCCTCTACTCTCTACTCCATCTAGTGGTCTGGTCATCACCATCACCATCCGCGGCTCTTGAAGTCTCTTTTCAGGACCCGGGTAC
 AsnAsnAsnProProThrProThrProSerSerGlySerGlyHisHisHisHisHisHisSerAlaAlaLeuGluValLeuPheGlnGlyProGlyTyr

BamH I
 EcoR I
 BsrG I
 Stu I
 Asc I
 PshA I
 Aat II
 Sal I
 Hind III
 Eag I
 Not I
 EcoCR I
 S*Tag
 CAGGATCCGAATTCGTACAGGCTTGGCGCGCCGACGTCGTCACAAGCTTGGCGCGCAGAGCTCGCTCTGGTGCCACGCGGTAGTAAAGAAACC
 GlnAspProAsnSerValGlnAlaLeuAlaArgProThrSerValAspLysLeuAlaAlaAlaGluLeuAlaLeuValProArgGlySerLysGluThr

S*Tag
 Xho I
 Pac I
 Avr II
 GCTGCTGCTAAATTCGAACGCCAGCAGATGGACAGCTCTACTTCTGCTGCTCTCGAGGCTTAATTAACCTAGGCTGCTAAACAAGCCGGAAGGAAGC
 AlaAlaAlaLysPheGluArgGlnHisMetAspSerSerThrSerAlaAlaLeuGluAlaEnd
 AS S*Tag 18mer Primer
 #71262-3

pET-49b(+) cloning/expresson region

A 2-2 Protein sequences of all human CARDs. These CARD constructs have been used for the small scale expression and solubility screening in Chapter 3.

Protein	Residues	Primary Sequence
NOD1	9-104	MEIIPSESHPHIQLLKSNRELLVTHIRNTQCLVDNLLKNDYFSAEDA EIVCACPTQPDKVRKILDLVQSKGEEVSEFFLYLLQQLADAYVDLRP WL
NOD1 ext	15-138	ESHPHIQLLKSNRELLVTHIRNTQCLVDNLLKNDYFSAEDAEIVCAC PTQPDKVRKILDLVQSKGEEVSEFFLYLLQQLADAYVDLRPWLLEIG FSPSLLTQSKVVVNTDPVSRYTQQLRHHLG
NOD2 ab	28-218	MCSQEAFFQAQRSQVLELLVSGSLEGFESVLDWLLSWEVLSWEDYEGF HLLGQPLSHLARRLLDTVWNKGTWACQKLIAAAQEAQADSQSPKLHG CWDPHSLHPARDLQSHRPAIVRRLHSHVENMLDLAWERGFVSQYECD EIRLPIFTPSQRARRLLDLATVKANGLAAFLQHVQELPVPLALPLE AAT
NOD2 a	28-120	MCSQEAFFQAQRSQVLELLVSGSLEGFESVLDWLLSWEVLSWEDYEGF HLLGQPLSHLARRLLDTVWNKGTWACQKLIAAAQEAQADSQSPKLH
NOD2 b	120-218	HGCWDPHSLHPARDLQSHRPAIVRRLHSHVENMLDLAWERGFVSQYE CDEIRLPIFTPSQRARRLLDLATVKANGLAAFLQHVQELPVPLALP LEAAT
RIP2	435-528	IAQQWIQSKREDIVNQMTACLNQSLDALLSRDLIMKEDYELVSTKP TRTSKVRQLLDTTDIQGEFAKVIVQKLKDNKQMGQPYPEILVSR
IPAF	1-92	MNFIKDNSRALIQRMGMTVIKQITDDLFWNVNLNREEVNIICCEKVE QDAARGIIHMLKKGSESCNLFKSLKEWNYPLFQDLNGQSLFHQ
NALP1	1377-1469	QLLHFVDQYREQLIARVTSVEVVLDKLGQVLSQEQYERVLAENTRP SQMRKLFSLSQSWDRKCKDGLYQALKETHPHLIMELWEKGSKKG
Caspase-1	1-92	MADKVLKEKRKLFIRSMGEGTINGLLDELLQTRVLNKEEMEKVKREN ATVMDKTRALIDSVIPKGAQACQICITYICEEDSYLAGTLGLSAD
ASC	105-195	QSAAKPGLHFIDQHRAALIARVTNVEWLLDALYGKVLTDQYQAVRA EPTNPSKMRKLFSTPAWNWTCKDLLLLQALRESQSYLVEDLERS
Caspase-9	1-112	MDEADRRLRLRRCLRLVEELQVDQLWDALLSRELFPHMIEDIQRAG SGSRRDQARQLIIDLETRGSQALPLFISCLEDTGQDMLASFLRTNRQ AAKLSKPTLENLTPVVLR
APAF1	1-98	MDAKARNCLLQHREALEKDIKTSYIMDHMISDGFLTISEEEKVRNEP TQQQRAAMLIKMILKKDNDYSYVSFYNALLHEGYKDLAALLHDGIPVV SSSS

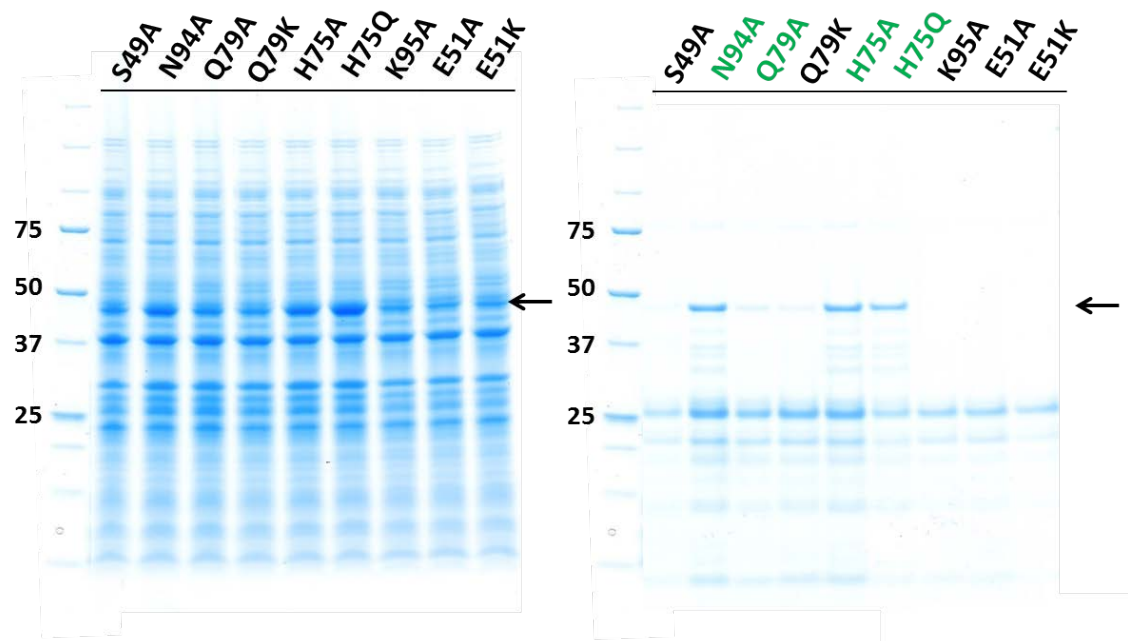
Caspase-2	28-128	GVCGMHPHHQETLKKNRVVLAKQLLLSELLEHLLEKDIITLEMRELI QAKVGSFSQNVELLNLLPKRGPQAFDAFCEALRETKQGHLEDMLLTT LSGLQHV
CARD8	336-431	ASAPPPFSGAAVVKENHRQLQARMGDLKGVLDLQDNEVLTENEKEL VEQEKTRQSKNEALLSMVEKKGDALDVLFRSISERDPYLVSYLRQQ NL
CARD9	1-104	MSDYENDEECWSVLEGRVTLTSVIDPSRITPYLRQCKVLNPDDEEQ VLSDPNLVIRKRKVGVLDDILQRTGHKGYVAFLESLELYYPQLYKKV TGKEPARVFS
RAIDD	1-101	MEARDKQVLRSLRLELGAEVLVEGLVLQYLYQEGILTENHIQEINAQ TTGLRKTMLLLLDILPSRGPKAFDTFLDSLQEFPPWREKLKKAREEAM TDLPGAD
RAIDD ext	1-136	MEARDKQVLRSLRLELGAEVLVEGLVLQYLYQEGILTENHIQEINAQ TTGLRKTMLLLLDILPSRGPKAFDTFLDSLQEFPPWREKLKKAREEAM TDLPGDRLTGIPSHILNSSPSDRQINQLAQRGPEWEPMVL
Iceberg	1-90	MADQLLRKKRRIFIHSGAGTINALLDCLLEDEVISQEDMNKVRDEN DTVMDKARVLIDLVTGKGPKSCCKFIKHLCEEDPQLASKMGLH
RIG-I ab	2-190	TTEQRRSLQAFQDYIRKTLDPITYILSYMAPWFREEEVQYIQAENNK GPMEAATLFLKFLLELQEEGWFRGFLDALDHAGYSGLYEAIESWDFK KIEKLEEYRLLKRLQPEFKTRIIPTDIISDLSECLINQECEILQI CSTKGMMAGAEKLVECLLRSDKENWPCTLKLALEKERNKFSELWIVE K
RIG-I a	2-91	TTEQRRSLQAFQDYIRKTLDPITYILSYMAPWFREEEVQYIQAENNK GPMEAATLFLKFLLELQEEGWFRGFLDALDHAGYSGLYEAIES
RIG-I b	91-190	SWDFKKIEKLEEYRLLKRLQPEFKTRIIPTDIISDLSECLINQECE EILQICSTKGMMAGAEKLVECLLRSDKENWPCTLKLALEKERNKFSE LWIVEK
MDA5 ab	1-201	MSGYSTDENFRYLISCFRARVKMYIQVEPVLDTLFLPAEVKEQIQ RTVATSGNMQAVELLLSTLEKGVWHLGWTREFVEALRRTGSPLAARY MNPCLTDLPSPSFENAHDEYLQLLNLLQPTLVDKLLVRDVLDKCME ELLTIEDNRNRIAAENNGNESGVRELLKRIVQKENWFSFNLNVRQT GNNELVQELTGSD
MDA5 a	1-100	MSGYSTDENFRYLISCFRARVKMYIQVEPVLDTLFLPAEVKEQIQ RTVATSGNMQAVELLLSTLEKGVWHLGWTREFVEALRRTGSPLAARY MNPCLT
MDA5 b	112-201	DEYLQLLNLLQPTLVDKLLVRDVLDKCMEELLTIEDNRNRIAAENN GNESGVRELLKRIVQKENWFSFNLNVRQTGNNELVQELTGSD
MAVS	1-95	MPFAEDKTYKYICRNFSNFCNVDVVEILPYLPCLTARDQDRLRATCT LSGNRDTLWHLFNTLQRRPGWVEYFIAALRGCELVDLADEVASVYQS Y
Caspase-4	1-95	MAEGNHRKKPLKVLES LGKDFLTGVLDNLVEQNVLNWKEEEKKKYYD AKTEDKVRVMADSMQEQRMAGQMLLQTTFFNIDQISPNKKAHPNMEA

Caspase-5	50-150	PNTDQKSTSVKKDNHKKKTVMKLEFYLGKDVLHG VFNYLAKHDVLT LK EEEEKKYYDTKIEDKALILVDSL RKNRVAHQMFTQTLLNMDQKITSV KPLLQIE
Caspase-12	1-95	MADEKPSNGVLVHVMVKLLIKTFLDGIFDDL MENNV LNTDEIHLIGKC LK FVV SNAENLVDDITETAQTAGKIFREHLWNSKKQLSSDISSD
BCL10	10-114	PSLTEEDLTEVKKDALENLRVYLCEKIIAERHFDHLRAKKILSREDT EEIS CRTSSRK RAGKLLDYLQENPKGLDTLVESIRREKTQNFLIQKI TDEV LKLRNIKLEHL
CARMA1	18-110	EEDALWENVECNRHMLSRYINPAKLTPYLRQCKVIDEQDEDEV L NAP MLPSKINRAGRLLDILHTKGQRGYVVFLESLEFYYPELYKLV TGKE
CARMA2	6-111	RRDSALTALDEETLWEMMESHRHRIVRCICPSRLTPYLRQAKVLCQL DEEEVLHSPRLTNSAMRAGHLLDLLKTRGKNGAIAFLES LKFHNPDV YTLVTGLQPDVD
CARMA3	12-118	EEAGAGSGSEAEEDALWERIEGVRHRLARALNPAKLTPYLRQCRVID EQDEEEVLSTYRFPCRVNRTGR LMDILRCRGKRGYEAFL EALEFYYP EHFTLLTGQEP AQ
CARD6	1-103	MATESTPSEIIERERKKLLEILQHDPDSILDTLTSRRLISEEEYETL ENVTDLLKKSRLKLLILVQKKGEATCQHFLKCLFSTFPQSAAICGLRH EVLKHENTV
COP1	1-99	MADKVLKEKRKLFIHSMGEGTINGLLDELLQTRVLNQEEMEKVKREN ATVMDKTRALIDSVIPKGAQACQICITYICEEDSYLAETLGLS AALQ AVQDN
ARC	1-95	MGNAQERPSETIDRERKRLVETLQADSGLLLDALLARGVLTGP EYEA LDALPDAERRVRRL LLLVQ GKGEAACQELLRCAQRTAGAPDPAWDWQ H
BinCARD	1-116	MTDQTYCDRLVQDTPFLTGHGR LSEQQVDRIILQLNRYYPQILTNKE AEKFRNPKASLRVRLCDLLSHLQRSGERDCQEFYRALYIHAQPLHSR LPSRHALRK FHITNHACLV LAR

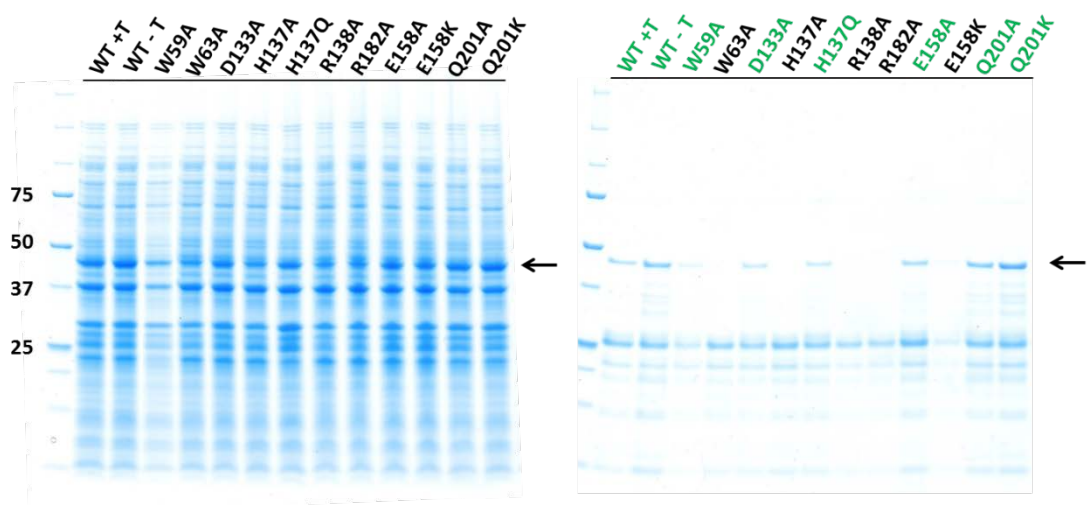
A 2-3 Protein sequences of protein-tags.

Tag	aa	M _w [kDa]	Primary Sequence
His	20	2.2	MAHHHHHHHSAALEVLFQ' <u>GPG...</u>
GST	237	27.5	MAPILGYWKIKGLVQPTRLLEYLEEKYEEHLYERDEGDK WRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHN MLGGCPKERAIEISMLEGAVLDIRYGVSR IAYSKDFETLKV DFLSKLP EMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALD VVLYMDPMCLDAFPKLVCFKKRIEAI PQIDKYLKSSKYIA WPLQGWQATFGGGDHPPKSDLVPRHNQALEVLFQ' <u>GPG...</u>
His-MBP	392	43.0	MAHHHHHHHSAAKIEEGKLVIWINGDKGYNGLAIEVGKKFEK DTGIKVTVEHPDKLEEKFPQVAATGDGPDIIIFWAHDRFGG YAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIA VEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMF NLQEPYFTWPLIAADGGYAFKYENGYDIKDVGVNDAGAK AGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGP WAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAA SPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEE ELVKDPRIAATMENAQKGEIMPNI PQMSAFWYAVRTAVIN AASGRQTVDEALKDAQTNSSSAALEVLFQ' <u>GPG...</u>

A 4-1 Small scale expression and solubility screening of NOD2 tandem CARDs with mutations inserted in CARDa.



A 4-2 Small scale expression and solubility screening of NOD2 tandem CARDs with mutations inserted in CARDb. +T: plus thrombin cleavage site, -T: original wildtype linker.



A 6-1 Sequence alignment of NOD2-CARDs with tandem DEDs. The primary sequences of Caspase-8 and MC159 DEDs are aligned with the two CARDs of NOD2. Highlighted in green are residues involved in the DED intramolecular interaction. Additionally, in MC159, highlighted residues are distinguished between fully buried (green) and partially buried (light green). S49, W59, W693 of NOD2-CARDa are shown in red, whereas other residues analysed via ITC but not affecting CARDb or CARDa binding respectively are shown in grey. Alignment was performed using ClustalW within the JalView application.

